

# THIS WEEK

## EDITORIALS

**DATA** US public health-care project needs to learn from UK mistakes **p.432**

**WORLD VIEW** Scientific diplomacy with Iran must respect human rights **p.433**

**MIXED UP** Brazilian bug is first female insect found with a penis **p.435**



## A toxic legacy

*Illegal dumping of toxic waste in the Italian Campania has been blamed for high rates of ill health in the region. The suspected link needs to be investigated using the most modern methods.*

It is tempting to see health hazards everywhere, especially if one believes the tabloid press. From the air we breathe to the food we eat and the fun we have, the risks of normal, everyday life have never received more attention.

But the people who live in the fertile agricultural landscape of Italy's Campania region, around the historic city of Naples, have genuine reason to be anxious about their health. For decades they have lived on top of potentially lethal toxic waste, illegally and secretly dumped there by the mafia. Rates of some cancers are higher, and lifespans shorter, than elsewhere in Italy. Complaints, though, have been continually silenced.

But Italian environmental police, working closely with anti-mafia forces, have used smart technologies to discover the subterranean dumps and are now systematically excavating them. Environmental officials are analysing soil and water. The crater walls look like geological strata, each layer formed by different types of waste: asbestos from demolished buildings, dioxin-rich chemical sludge, drums of solvent, the odd motor vehicle. Black water forms pools at the bottom. No one is allowed to even look into the stinking, steaming pits without a respirator mask.

Massive street demonstrations in Naples last autumn prompted officials to take action. The national government approved funding earlier this year for a two-year programme in which the several hundred thousand people who live close to the dumps will be screened for cancers that have been linked to environmental exposure. The health ministry has analysed all the health and environment studies carried out so far and concluded that there is no evidence yet to link the dumps to cancer. Yet local people are convinced that noxious run-off from the dumps is gradually killing them and their families, loading their bodies with toxic chemicals and twisting their cells into tumours. They want answers, and they want the scientists to supply them.

### TOXIC BLACK BOX

Environmental police have so far identified 32 sites containing an estimated minimum of 3.5 million cubic metres of toxic waste. But without crucial information about actual exposure, including dose levels, it is impossible to determine whether the dumped chemicals have raised cancer risk in what is a poor region, where people smoke more and have healthier lifestyles than in other parts of the country.

Similar issues arise whenever cancer clusters emerge around nuclear power plants or industrial sites. Attempts to prove a causal link face several dilemmas. One is that the number of cancer cases is usually too small for conclusive statistics. Another is that the cancers usually become apparent years after the hypothetical exposure to carcinogen, and such historical exposures are almost impossible to prove scientifically. But usually, the identity of the feared chemicals is known.

In Campania, the challenge is greater because the very fundamentals are unknown — not only the location and chemical content of the

dumps, but also the true local incidence of cancer. Over the past year or so, that information has started to emerge. And biologists there are wondering if they might be able to tie their whole noxious experience into the global research effort to develop methods to prove chemical exposure scientifically. Could their poisoned fields serve as a giant experiment in the new science of 'exposomics', which aims to identify biomarkers of past and present exposure to toxic environmental chemicals?

"Campania could be a perfect field study for a biomonitoring research programme," says Gennaro Ciliberto, scientific director of the National Tumour Institute Pascale in Naples.

**"Campania could be a perfect field study for biomonitoring research."**

Precedents do exist. One is in the city of Thessaloniki in northern Greece, where newly impoverished people have started to burn more biomass for home heating, contributing to smog. A research project funded by the European Union combines regular analyses of atmospheric pollutants with analyses of urine and blood of a cohort of

people to quantitatively determine how their gene-expression, protein and metabolite profiles change as a result of pollutants entering their bodies.

Italy, in austerity mode, has little money to spare for research. So Ciliberto has a suggestion. "This is the sort of programme that should be funded from the structural funds that the European Commission has awarded the region."

He could be right. These subsidies came to a total of €6.9 billion (US\$9.6 billion) for Campania alone in 2007–13, and the commission explicitly encourages use of the funds to boost local research and development capacity. Allocations of the next tranche of structural funds within Italy (2014–20) are currently under discussion. Ciliberto's idea deserves serious consideration.

The ancient Romans called the area *Campania felix*, or fortunate countryside, because of the soil's fertility (conferred by its frequent coating with volcanic ash from Mount Vesuvius's deadly eruptions). But modern Italians call it *Campania infelix*, and its agriculture-based economy is suffering from public perception that its products could be contaminated.

Anti-mafia action made the waste-disposal racket too public to continue after 2003. But mafia domination continued to stifle public discussion about the environmental poisoning. Locals, government officials and even many academics have been reluctant to talk. Some feared retribution, but more feared that they or their friends would be exposed as having helped in the logistics of the operations, or as having accepted mafia bribes for turning a blind eye. People began setting fire to the dumps, making matters worse.

The fires are out now, but the questions of ill health remain. They should be properly investigated. Even if the results prove inconclusive, the information gained would be worth it. ■

# A question of trust

*NASA's decision to renege on SOFIA project casts doubts over its reliability as a partner.*

For sale: one flying observatory, barely used. Price: US\$1 million per night. That is essentially the advertisement just posted by NASA in a bid to save its Stratospheric Observatory for Infrared Astronomy (SOFIA) — a jumbo jet fitted with a 2.5-metre infrared telescope. The plea for cash comes in the wake of the space agency's surprise announcement last month that it would no longer pay the roughly \$80 million a year it had planned to spend on SOFIA.

The cancellation was a shock because SOFIA had only just reached full operational capacity (see *Nature* 506, 420–421; 2014). After more than a decade in development, and a bill of \$1.25 billion, the plane finally had a full suite of astronomical instruments in place. Plans were being laid for SOFIA to start increasing its flight schedule from its home in Palmdale, California, towards the goal of being airborne from dusk to dawn several nights a week, gathering science on key astronomical questions. Too little too late, NASA said — years of delays mean that SOFIA has missed its chance to overlap with other infrared space observatories and, the agency claimed, provide sufficient science return.

But there is a deeper issue: NASA is not the only participant in SOFIA. From the project's start, the German aerospace centre DLR has paid one-fifth of the cost, granting German scientists access to 20% of its observing time. When NASA pulled out, the DLR was left hanging high and dry.

What will happen next is far from certain, and will be dictated by complex financial negotiations between various branches of the US government. If NASA sticks to its plan, project managers must wind down all science flights and mothball SOFIA by 30 September, the end of the current fiscal year. But the \$12 million currently budgeted is not enough to do even that job, according to a 15 April report from the Government Accountability Office.

It is possible that the US Congress will restore some amount of money to the project, overriding NASA's request. Intense lobbying is now under way to sway the minds of key members of Congress, such as those who represent SOFIA workers in California. This may well succeed, granting the project some fraction of NASA's original budget, which could be enough to keep it limping along on a reduced schedule and with fewer staff.

A third option is that some as-yet-unknown partner might emerge from the shadows to pick up some of NASA's \$80-million-a-year share.

**"NASA must expect other countries to ask some tough questions before they sign up to cooperative projects."**

Hence the offer to rent it out. *Nature* would be delighted to be proved wrong, but demand for that seems unlikely. Germany, for one, has already told NASA it cannot pay more than it is already.

What happens to SOFIA has ramifications far beyond the world of infrared astronomy. It has major implications for any country that might want to share scientific projects with NASA, and for the responsibility that NASA

has to honour those agreements. The deal between NASA and the DLR contains a clause that allows one partner to propose shuttering the observatory. German officials and scientists were understandably furious at NASA's unilateral decision to shut it down.

During its long and chequered history, SOFIA has come close to being cancelled before. And there are many valid reasons to question its scientific return per dollar spent — or whether it should be flying at all (see *Nature* 466, 413; 2010). But what is happening with SOFIA is circumventing the usual process for making hard decisions in tight fiscal times. NASA has 'decadal surveys' to set priorities for future missions, and 'senior reviews' to check the worth of operational ones and to halt those that do not make the cut.

For the moment, SOFIA continues to fly as normal. A NASA/DLR working group is outlining options for how to wind down the programme; its report is due by the end of April, and is likely to read like the order of service for a funeral.

No matter whether Congress — or some benefactor — steps in to help, the breach of trust with the DLR is irrevocable. NASA must expect other countries to ask some tough questions before they sign up to cooperative projects in the future. ■

## Take care

*The United States must tread carefully when building a health-data system.*

With last year's revelations about the National Security Agency's intrusive spying practices, and the disastrous roll-out of the US health-insurance website HealthCare.gov, now may not seem the best time for the United States to establish a national network to collect and analyse health data from millions of patients.

Good luck, then, to the government-backed Patient-Centered Outcomes Research Institute (PCORI), which is trying to do just that.

In December, the institute in Washington DC launched the PCORnet project to collate medical records, physiological data and insurance claims from as many as 30 million US patients through 29 web networks — a dream cohort for any biomedical researcher. Last month, the organization began building the computer infrastructure that will tie these records together, allowing researchers to spot trends in how lifestyles affect health, which drugs are most effective for certain types of individual, and so on. If it works as advertised, these findings could save millions of dollars by allowing researchers to recruit the right participants for clinical trials and

doctors to prescribe the right treatments for individuals.

Designing a US\$94-million programme that draws big data from dozens of computer systems is an enormous challenge — just ask former Health and Human Services secretary Kathleen Sebelius, who resigned last week after months of technical problems with HealthCare.gov. But the ethical issues may prove an even higher hurdle. The United States would do well to watch the United Kingdom as it grapples with creating its own national database, care.data (see *Nature* 507, 7; 2014). Although the UK national health system is much more mature and cohesive than anything the United States has, the project's rollout has been delayed for months because of unanswered questions about informed consent and access to sensitive patient data.

Even if PCORI can solve its technical and ethical problems, it will be some time before the project measures up to systems such as Denmark's MedCom, which collates comprehensive patient data going back to 1977 and provides an indispensable resource for researchers doing longitudinal studies. The system seems to work, largely because of its transparency about how the data are used.

PCORI says that it will let patients help to determine research priorities, and promises that only anonymous data and no individual records will be available to researchers. It is crucial that the institute follows through, to maintain patients' trust. Without that, neither researchers nor patients will benefit. As the old saying goes, you make your own luck. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)





## Iran needs to present a united front on science

*Iranian scientists must be allowed to interact abroad without fear of persecution, says Elise Auerbach.*

Iranian officials last month unveiled the Mustafa Science and Technology prizes, for researchers in Islamic countries. The creation of the four US\$500,000 prizes illustrates the growing importance to the Iranian leadership of domestic science and the nurturing of scientific cooperation and exchange with other nations.

The Iranian government has already developed a cadre of distinguished scientists. State-run media routinely vaunt advances made by Iranian researchers, often new military systems. Iran hosted the Fifth International Conference of Cognitive Science in Tehran last year. And Hessamaddin Arfaei, a physicist at the Institute for Research in Fundamental Sciences in Tehran, visited the United States in April 2005 to deliver a lecture at a meeting of the American Physical Society and to continue his efforts “to establish interaction between Iranian scientists and the international scientific community”. His trip was presumably carried out with the blessing of Iranian officials.

But these efforts are being subverted by other sectors of Iran’s ruling establishment, which are determined to ferret out ‘traitors’ among the country’s most promising scientists. As Iran’s interest in science and demand for international cooperation grows, researchers abroad have an opportunity to tip the balance of power away from these hardliners.

Too often, scientists in Iran are victims of a fantasy that there is a massive campaign orchestrated by the US government (including the CIA), Israel’s Mossad and the British intelligence services to launch a ‘Velvet Revolution’ (*Enghelab-e Makhmal*) to undermine the foundation of the Islamic Republic of Iran. To achieve their aim, the forces behind the conspiracy are alleged to be recruiting high-profile Iranians through academic exchanges, by attracting them to attend conferences abroad or by inviting them to hold residency at Western institutions.

Hardline Iranian newspapers and websites recite every known or imagined contact between Iranian scholars and their European and US colleagues, no matter how benign, including research grants and visiting fellowships at universities and private foundations. The conspiracy about the Velvet Revolution, and the concomitant requirement to seek out and uncover ‘enemies’ — the higher profile the better — serves only the interests of the security agencies and the hardliners, who can justify the power and scope of the elaborate (and costly) security apparatus by claiming its necessity as a bulwark against those aiming to destroy the country.

The physicist Omid Kokabee, for example, was arrested on spurious charges of crimes against national security in 2011 and given a ten-year prison sentence in 2012. Kokabee, who was pursuing a PhD in quantum optics at the University of Texas at Austin, had resisted repeated invitations from

Iranian authorities to work on military and intelligence projects. His claimed crimes were communicating with a hostile government and receiving illegitimate funds — an apparent reference to the normal stipend given to graduate students at his university.

His case and others like it must have a chilling effect on other Iranian scientists engaged in legitimate interactions with colleagues in the West. This inhibits the advancement of science in Iran and seems to be utterly counterproductive to the state-endorsed efforts to promote Iran’s scientific achievements internationally.

The more moderate members of Iran’s ruling establishment surely understand that the country’s successful integration into the global community is undermined if its best and brightest scholars fear the consequences of engaging in normal interactions with colleagues

abroad. The reality of Iran’s ‘brain drain’ problem was recognized by its science and technology minister, Reza Faraji Dana, who noted in January: “Every year, about 150,000 of our elite emigrate from Iran, costing our economy \$150 billion.”

The harsh penalties imposed on Kokabee and others have earned Iran the condemnation not only of the human-rights organizations campaigning on his behalf, but also of scientific and scholarly societies — including the American Physical Society, with which the country is otherwise engaged in a dialogue to improve opportunities for collaboration.

Scientific diplomacy can help to build bridges and defuse international tension. In an article titled ‘Which Iran Will We Choose?’, published in December by the *The WorldPost*, academics Trita Parsi, Bijan Khajepour and Reza Marashi argued that sustained engagement with Iran on science projects

“can spill over to other crucial areas, such as the human rights situation”.

A necessary precondition for such collaboration must be the right of scientists to engage in their normal professional activities without incurring persecution. Iran’s scientists, its population and the broader scientific community would all benefit from Kokabee’s release. Campaigns to free Iranian scientists persecuted in this way have succeeded in the past. Brothers Arash and Kamiar Alaei, both HIV/AIDS physicians, were arrested in 2008 and convicted of “communicating with an enemy government”. They were released following international pressure.

By offering the prospect of shared scientific enterprises on one hand, and advocating for an end to the persecution of scholars on the other, science diplomacy could help to convince decision-makers in Iran that the benefits of international scientific cooperation greatly outweigh any value in placating those who prefer to perpetuate a giant conspiracy. ■

SCIENTIFIC  
DIPLOMACY CAN  
HELP TO  
**BUILD  
BRIDGES**  
AND DEFUSE  
INTERNATIONAL  
**TENSION.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/sngks4](http://go.nature.com/sngks4)

**Elise Auerbach** is the Iran country specialist with Amnesty International USA in Chicago, Illinois.  
e-mail: [aiis@uchicago.edu](mailto:aiis@uchicago.edu)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## GEOLOGY

### Ancient impact recorded in rock

South African rocks reveal that a huge asteroid smashed into Earth 3.26 billion years ago, turning sediments to liquid.

The rocks in the Barberton greenstone belt in South Africa contain tiny blobs of material that condensed from clouds of rock vapour generated by ancient impacts. Norman Sleep and Donald Lowe of Stanford University in California used those blobs, along with other features of the Barberton rocks, to calculate the size and geological effects of this particular asteroid impact.

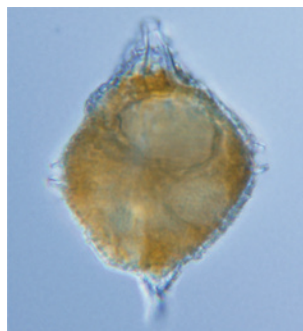
The duo found that the asteroid was probably 50 kilometres across. Seismic waves from the impact fractured much of Earth's crust, triggering strong earthquakes and giant tsunamis, the authors say. *Geochim. Geophys. Geosys.* <http://doi.org/sb4> (2014)

## ORGANISMAL BIOLOGY

### 'Extinct' plankton found in Pacific

A plankton species thought to have gone extinct more than a million years ago has been found living in the western Pacific Ocean.

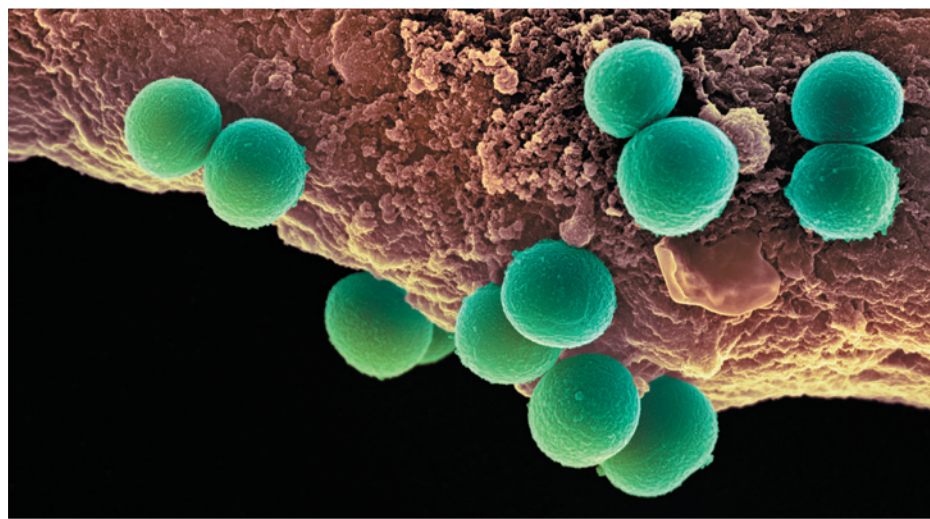
Kenneth Mertens of Ghent



University in Belgium and his colleagues discovered live cysts of the dinoflagellate plankton *Dapsilidinium pastielsii* (pictured) in the sea floor from Japan to the Philippines and Indonesia. This stably warm region, known as the Indo-Pacific Warm Pool, harbours high levels of biodiversity.

The warm waters there could have provided a haven for *D. pastielsii* to survive in the water column and sediment. Such areas of refuge will be important for maintaining biodiversity in a changing climate, the authors say.

*Geology* <http://doi.org/sb7> (2014)



## ANTIMICROBIALS

### Biocide boosts bacterial binding

People exposed to an antimicrobial compound used in many consumer products are more likely to carry an opportunistic pathogen, increasing their risk of infection.

The bacterium *Staphylococcus aureus* (pictured in green) is normally harmless in healthy adults but can cause serious infections in sick or injured people. A team led by Blaise Boles at the University of Michigan in Ann Arbor found that people with higher levels of the biocide triclosan in their nasal secretions

had *S. aureus* colonies in their noses more frequently than people with little or no triclosan in their secretions. Triclosan, which is used in toothpastes, soaps and medical equipment, increased the binding of *S. aureus* to plastic, glass and human proteins such as collagen and keratin.

The results suggest an urgent need to re-evaluate the use of triclosan in consumer products, the authors say.

*mBio* 5, e01015-13 (2014)

## ANIMAL BEHAVIOUR

### Fruit-fly microbes draw more flies

Researchers have pinpointed a chemical source that attracts fruit flies to food: the gut bacteria of fruit-fly larvae sitting on the food.

Reuven Dukas and his colleagues at McMaster University in Hamilton, Canada, previously showed that fruit flies (*Drosophila melanogaster*) are attracted to food occupied by larvae. Now they find that the flies prefer food that has been inhabited by larvae with intact gut

microbiomes rather than by bacteria-free larvae.

The authors also show that this attraction is not related to the presence of gut bacteria or larvae in the food, but to the physical changes that the feeding larvae make. Used food is easier for a larva to burrow into than fresh food, they say. *J. Exp. Biol.* 217, 1346-1352 (2014)

## ASTRONOMY

### Earth twin spotted in habitable zone

Astronomers have discovered a planet that is a similar size to Earth orbiting a cool, dim

SPL

YOSHIMITO TAKANO

star at just the right distance for liquid water to exist. It is the most Earth-like planet found so far that could potentially host life.

A team led by Elisa Quintana of the SETI Institute in Mountain View, California, discovered the planet, Kepler-186f, by looking at data from NASA's Kepler space telescope. Slight drops in a star's brightness revealed the orbiting planet, the fifth known in that particular system.

Kepler-186f is just 1.1 times the size of Earth and orbits its star every 130 days. It is towards the outer edge of the star's habitable zone — in our Solar System, it would be equivalent to putting Earth out near the orbit of Mars.

*Science* 344, 277–280 (2014)  
For a longer story on this research, see [go.nature.com/tmbgje](http://go.nature.com/tmbgje)

## MICROBIAL EVOLUTION

## How a flesh-eater evolved

A deadly wave of flesh-eating bacteria emerged in the 1980s when a single cell gained a pair of toxin genes from another bacterium.

Infections with group A *Streptococcus* (GAS) strains, which cause necrotizing fasciitis, spiked worldwide in the late 1980s and early 1990s. To piece together the sequence of molecular events behind this epidemic, James Musser at the Houston Methodist Research Institute in Texas and his team sequenced the genomes of 3,615 GAS strains isolated from Europe and North America between 1969 and 2013.

Analysis of the genomes revealed a series of genetic changes that began to occur in the early 1970s. The final changes that took place before the strain became a voracious flesh-eater happened in around 1983, when less-virulent strains acquired the genes that are needed to produce certain toxins in humans.

*Proc. Natl Acad. Sci. USA*  
<http://doi.org/sdc> (2014)

## PHYSICS

## Liquid bubbles stop sound

Liquid foam can block the transmission of ultrasound, report Valentin Leroy at Paris Diderot University and his colleagues.

The French team created the foam by bubbling an insoluble gas, perfluorohexane, into water containing a surfactant, and then sent ultrasound through the foam.

The elastic, thin films at the edges of adjacent foam bubbles reflected the sound waves. The exact frequencies that were blocked varied with the average bubble size, the biggest of which measured up to 50 micrometres in radius.

The findings show that foams can act as acoustic metamaterials — materials that block sound waves because of their unique structure.

*Phys. Rev. Lett.* 112, 148307 (2014)

## MICROBIOLOGY

## Bacteria elbow out the competition

Some microorganisms push others out of the way to gain better access to oxygen.

Wook Kim and Kevin Foster at the University of Oxford, UK, studied the soil bacterium *Pseudomonas fluorescens*, which grows in dense layers called biofilms.

They found that a strain that consistently outgrew others and dominated the colony carried a mutation in a gene called *rsmE*. This mutant strain secretes polymers that expand, nudging neighbours aside and allowing cells to reach the colony surface where they can receive the most oxygen.

Life in tightly packed communities has selected for this bacterial pushiness, the authors say.

*Proc. Natl Acad. Sci. USA*  
<http://doi.org/sb5> (2014)

## SOCIAL SELECTION

Popular articles  
on social media

## Illegal fishing hooks online attention

Researchers on Twitter were astounded by some of the figures given in a paper describing extensive illegal fish imports into the United States. According to Tony Pitcher at the University of British Columbia in Vancouver and his colleagues, in 2011 more than 20% of wild-caught seafood coming into the country — worth between US\$1.3 billion and \$2.1 billion — was from illegal or unreported fishing. To make their estimates, the researchers used data from the US National Marine Fisheries Service for 30 major seafood products from 10 countries. Illegal, unreported and unregulated fishing disrupts ecosystems, food security and livelihoods around the world, the authors say.

*Mar. Policy* 48, 102–113 (2014)



Based on data from altmetric.com.  
Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/o5v8cv](http://go.nature.com/o5v8cv)



## ZOOLOGY

## Sex changed in Brazilian insect

Female insects in Brazilian caves initiate reproduction, and they do so with a penis.

Kazunori Yoshizawa at Hokkaido University in Sapporo, Japan, and his colleagues studied four *Neotrogla* species, one of which copulates for up to 70 hours. The authors found that the female uses an external organ (**pictured**) during sex to penetrate the male and to receive sperm for both reproduction and nutrition. All of the species have spines on the penis-like

structure, which the females use to anchor themselves in the male genital chamber.

The competition between females for mates might have driven the evolution of this elaborate sexual structure, the authors say. However, this is the only known case of reversed sex organs in animals with reversed sexual roles, they add.

*Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2014.03.022> (2014)

For a longer story on this research, see [go.nature.com/ptnetb](http://go.nature.com/ptnetb)

➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### NIH grant recycling

The US National Institutes of Health is to allow researchers to base new grant applications on ideas that have previously been rejected for funding. The agency announced the policy on 17 April ([go.nature.com/43cibf](http://go.nature.com/43cibf)). It modifies a 2009 rule that any completely new grant application must be substantially scientifically different from one that has been rejected. The change is in response to complaints that the 2009 policy penalized young researchers whose work cannot easily change direction to generate completely new proposals.

### Emissions down

US emissions of six major heat-trapping gases decreased by 3.4% between 2011 and 2012, to the equivalent of 6,526 million tonnes of carbon dioxide, the US Environmental Protection Agency said on 15 April. Energy consumption dropped in every sector of the US economy, the agency said, because more energy was generated from natural gas than from coal, among other factors. Emissions from transport also fell as fuel efficiency increased and demand for new vehicles declined.

### Environment boost

Some 30 nations pledged US\$4.43 billion on 16 April to help developing countries to protect the environment from degradation. The funding will be distributed over the next four years by the Global Environment Facility — an international partnership of governments, civil-society organizations and companies that supports sustainable development and environmental treaties such as the Convention on Biological



CHIP SOMODEVILLA/GETTY IMAGES

## T. rex comes to Washington

A near-complete 66-million-year-old *Tyrannosaurus rex* skeleton (right femur pictured) arrived at the US National Museum of Natural History in Washington DC on 15 April. It is on loan for 50 years from the US Army Corps of Engineers. The fossil is one of the largest and most complete *T. rex* specimens ever discovered, with up to 85% of the skeleton

recovered after it was found on federal land in 1988. It will become the centrepiece of the museum's new 2,900-square-metre dinosaur and fossil hall, which is scheduled to open in 2019 and will house examples from the museum's collection of 46 million fossils. The hall will be named after philanthropist David Koch, who donated US\$35 million to the project.

Diversity. The funding will go towards projects in more than 140 countries, aimed at tackling threats such as species extinction, toxic chemicals and deforestation.

### Cash for innovation

The European Parliament voted on 15 April in favour of a €22-billion (US\$30-billion) programme to fund research and innovation partnerships between academia and industry. The Innovation Investment Package is funded by the European Union, industry and member states. Much of it will support six public-private partnerships, revamping the Joint Technology Initiatives of Europe's previous

multi-year research-funding programme. The package will provide competitive grants in innovative medicines, aeronautics, bio-based industries, fuel cells and hydrogen, and electronics components. Member states are expected to approve it in time for a proposed July launch.

### Pollution survey

Just under one-sixth of China's land is polluted, according to the first nationwide soil survey, released on 17 April by the Ministry of Environmental Protection and the Ministry of Land and Resources. Much of the pollution comes from cadmium, nickel and arsenic released by industry and agriculture, the analysis found.

The survey ran from April 2005 to December 2013 and covered 6.3 million square kilometres — two-thirds of the nation's land area.

### GM-labels law

Vermont looks set to become the first US state to enact a law requiring the labelling of foods containing genetically modified (GM) ingredients. On 16 April, the state's senate passed a bill in favour of labelling. The house of representatives is expected to agree to amendments to the bill, which has already passed through the house and senate once. Vermont's governor supports GM labels. Connecticut and Maine passed similar laws last year, but both

contain clauses that stop them from coming into effect until neighbouring states follow suit.

## EVENTS

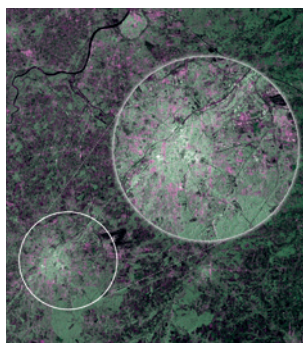
**LADEE crashes**

NASA's Moon-dust probe shattered into clouds of dust when it hit the far side of the Moon, as planned, in a mission-ending impact on 18 April. The Lunar Atmosphere and Dust Environment Explorer (LADEE) finished its duties in March. Mission controllers directed the crash to protect historic lunar locations such as the Apollo landing sites from probe debris. During its lifetime, LADEE made the best measurements yet of lunar dust, generated as tiny meteorites bombard the Moon's surface. See [go.nature.com/mwqqln](http://go.nature.com/mwqqln) for more.

**Weapons update**

Syria has removed or destroyed around 80% of its chemical-weapons materials, a joint mission of the Organisation for the Prohibition of Chemical Weapons and the United Nations announced on 19 April. The remaining materials should be removed over the next few days, the mission says; all materials must be destroyed by 30 June. The process began in January under an agreement brokered by Russia and the United States.

## RESEARCH

**Sentinel shots**

The European Space Agency published the first images from its Sentinel-1A satellite on 15 April. This image shows the city of Brussels and its surroundings. Green correlates to vegetation; pink to urban areas; white to high-density urban areas; and black to waterways and low-reflective areas such as airport runways. The image was captured by the satellite on 12 April, nine days after it launched.

**Egyptian satellite**

Egypt's second remote-sensing satellite launched on 16 April on a mission to collect agricultural, geological and environmental data from northern Africa. The Russian-built EgyptSat-2 sailed into space aboard a Soyuz launch vehicle from the Baikonur Cosmodrome in Kazakhstan. The probe replaces EgyptSat-1, which lasted

for three years after its 2007 launch. Data from the second satellite will be handled by the Egyptian National Authority for Remote Sensing and Space Sciences near Cairo.

## BUSINESS

**SpaceX delivery**

SpaceX, the private spaceflight company in Hawthorne, California, made its third successful cargo delivery to the International Space Station on 20 April, as part of a US\$1.6-billion contract with NASA to make a dozen drops to the space station by 2016. The flight was delayed from March because of technical difficulties. It carried science experiments and supplies including high-definition cameras for viewing Earth, and equipment to test why the immune system becomes suppressed in microgravity.

**Gene-cutter patent**

An acclaimed method for editing DNA sequences has won its first patent. On 15 April, the Broad Institute in Cambridge, Massachusetts, announced that it had patented an engineered CRISPR-Cas9 system. The 'clustered regularly interspaced short palindromic repeats' process uses an engineered bacterial enzyme to target genes, making precise cuts in their DNA sequence and thereby enabling them to

## COMING UP

**29 APRIL**

The first solar eclipse of the year takes place. A bright ring of sunlight will surround the dark Moon in this annular eclipse, which will be best viewed from Antarctica.

[go.nature.com/q3dvlq](http://go.nature.com/q3dvlq)

**30 APRIL**

The World Health Organization publishes a report on global surveillance of antimicrobial resistance. It will present data on antibiotic resistance in nine bacteria responsible for common maladies such as diarrhoea and bloodstream infections.

be modified. The system was devised by molecular biologist Feng Zhang at the Broad. Last November, Zhang co-founded a company, Editas Medicine in Cambridge, to develop ways of using the CRISPR technology for therapies (see [go.nature.com/c8jopb](http://go.nature.com/c8jopb)).

**RNAi drugs battle**

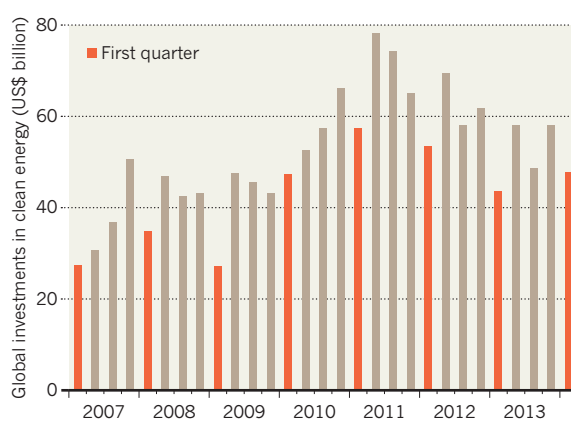
The withdrawal of the pharmaceutical firm Novartis from RNA interference (RNAi)-based drug development does not spell the end for the field, a biotechnology executive says. RNAi-based drugs use snippets of RNA to shut down the production of disease-causing proteins. In an interview with *Nature* on 15 April, John Maraganore, chief executive of RNAi specialists Alnylam in Cambridge, Massachusetts, said that the announcement on 14 April by Novartis of Basel, Switzerland, indicated that "large companies don't do a good job with the technology". The field is actually thriving, he says. See page 443 for more.

## TREND WATCH

Clean-energy analysts hope that a promising first quarter to 2014 foreshadow a rebound for dipping global investments. Figures released by Bloomberg New Energy Finance showed investment rising 10% on the same period last year, driven mainly by small-scale solar projects. There was a notable geographic shift too, with two of the largest deals occurring in Indonesia (US\$1.6 billion for the Sarulla geothermal project), and Kenya (\$860 million for the Lake Turkana wind project).

**HOPE FOR CLEAN-ENERGY INVESTMENT REBOUND**

Global investments reached US\$47.7 billion in the first quarter of 2014.





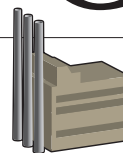
# NEWS IN FOCUS

**SEISMOLOGY** Chile earthquake redefines study of large quake zones **p.440**

**PHYSICS** Quantum networks gear up for real-world debut **p.441**

**BIOTECHNOLOGY** RNA interference shows fresh promise as treatment **p.443**

**ENERGY** The reactors that turn organic waste into fuel **p.448**



PICHI CHUANG/REUTERS/CORBIS



The drug Tamiflu is prescribed as the front-line treatment for serious cases of influenza.

## EPIDEMIOLOGY

# Tamiflu report comes under fire

*Conclusions on stockpiling of antiviral drugs challenged.*

BY DECLAN BUTLER

A study that calls into question the stockpiling of billions of dollars' worth of antiviral drugs to mitigate the threat of influenza pandemics has been criticized by flu researchers.

The analysis of Tamiflu and Relenza, drugs known as neuraminidase inhibitors, was published on 10 April by the Cochrane Collaboration<sup>1</sup>, a group that reviews the effectiveness of health-care measures. It concluded that the medicines were of little use. At the same time, the journal *BMJ* published a series of articles, including two that summarize the Cochrane findings<sup>2,3</sup>, and several editorials that focus on the five-year campaign by Cochrane and the *BMJ* to obtain the unpublished drug-company clinical-trial data later used in the review.

The results “challenge the historical assumption that neuraminidase inhibitors are effective

in combating influenza”, declared a joint *BMJ*–Cochrane news release on the findings. The drugs have had their “effectiveness overplayed, and harms underplayed”, said Fiona Godlee, the *BMJ*'s editor-in-chief, at a press conference. The study generated worldwide media coverage, including headlines labelling Tamiflu as “useless” and “ineffective”.

But the review and its bottom line are vigorously contested by many flu researchers. They argue that the analysis — an update by Cochrane — is based on randomized clinical trials (RCTs) of the drugs that lack sufficient statistical power to allow reliable conclusions to be drawn about the effects on flu complications and hospitalizations. These are the key outcomes of interest during a flu pandemic.

The critics say that the review also excluded many observational studies that have found the drugs to be helpful in normal clinical settings.

Tamiflu is prescribed as the main treatment

for serious cases of flu, and researchers worry that the media storm risks undermining public confidence in this class of drug. “We risk losing one of the few weapons we have, because of overly negative publicity,” says Peter Openshaw, director of the Centre for Respiratory Infection at Imperial College London.

The review by Cochrane, a non-profit organization in London, considered both Tamiflu (oseltamivir), produced by the drug company Roche in Basel, Switzerland, and Relenza (zanamivir), marketed by GlaxoSmithKline in Brentford, UK. It is one of the first analyses to review clinical-study reports — documents that companies submit to regulators and that contain fuller trial data than published RCTs.

Both drugs have been found to shorten the duration of flu-like symptoms, and to be effective in treating severe disease; what is in dispute is whether they reduce hospitalizations and severe complications. The Cochrane authors and the *BMJ* claim that there was no evidence for such gains. They also say that the findings bring into question the decision by governments in the mid-2000s to stockpile the drugs against the threat of a pandemic of the H5N1 avian-flu virus, which has a mortality rate of around 60%. Carl Heneghan of the University of Oxford, UK, a co-author of the review, said that there was “no credible way these drugs could prevent a pandemic” and that stockpiles were “money thrown down the drain”.

But a dozen experts contacted by *Nature*'s News team said that the clinical-study reports, although they allow more detailed analyses, offer no substantial fresh findings.

“The Cochrane authors have done a thorough review, and deserve credit for their efforts to obtain raw clinical data from Roche and regulatory authorities, but ultimately its findings are not surprising,” says Peter Horby, a flu researcher at the University of Oxford Clinical Research Unit in Hanoi, Vietnam.

Jody Lanard and Peter Sandman, independent risk-communication experts in New York, say that the press release on the review omits findings that in their opinion are key. For example, it rounds down Tamiflu's reported 17-hour reduction in symptom duration in adults to “just half a day”, and describes the reduction as “small”. It also does not report the 29-hour reduction in children. Lanard and Sandman claim that there has been “cherry-picking of the results to make them look worse for antivirals”.

Horby's view is that the *BMJ*–Cochrane ►



► press release and public statements have contributed to media misinterpretation of what the study does and does not say. “The review does not state the drugs are ‘ineffective’ or ‘useless,’” as was reported in some media stories, he says.

In a joint response to *Nature*, BMJ–Cochrane said that they felt the press release did “a good job of presenting the main findings”. They defend its presentation of the reduction in duration of symptoms as justified, and note that the fuller outcomes were included in the review’s summary of its findings. The press coverage, they say, was “not bad in terms of balance and accuracy”.

The two organizations agree that the limited statistical power of the trial data on the numbers of hospitalizations and complications makes it difficult to draw reliable conclusions on these. But they argue that their findings of no evidence for any effects challenge past analyses of trials that found that the drugs reduced complications and hospitalizations. They claim that the review did show new findings, for example data on side effects.

RCTs are considered the gold standard for establishing the effectiveness of drugs, and Cochrane restricts itself to meta-analyses of these in its review. But critics note that these small clinical trials were carried out to gain regulatory approval for Tamiflu as treatment and prophylaxis for seasonal flu, which in most cases is mild. This meant that the healthy trial subjects rarely developed complications. In other words, the trials were not designed to test for the severe outcomes that are most relevant to pandemics.

The critics add that observational studies of how large numbers of people respond to treatments under normal medical care can also provide important information on the effectiveness of drugs. In March, for example, an observational study<sup>1</sup> of 30,000 people hospitalized during the 2009–10 swine-flu pandemic reported that neuraminidase inhibitors reduced mortality by 25%.

But because of its policy of reviewing only RCTs, Cochrane did not include observational data. “Given the limited number of RCTs, and the considerable evidence base on the effectiveness of neuraminidase inhibitors over the past ten years or more, it is difficult to justify exclusion of the observational evidence,” says Ben Cowling, a flu epidemiologist at the University of Hong Kong.

BMJ–Cochrane said they exclude all observational studies because they are “unreliable for establishing treatment effects”. ■

1. Jefferson, T. et al. *Neuraminidase Inhibitors for Preventing and Treating Influenza in Healthy Adults and Children (Review)* (Wiley, 2014).
2. Jefferson, T. et al. *Br. Med. J.* **348**, g2545 (2014).
3. Heneghan, C. J. et al. *Br. Med. J.* **348**, g2547 (2014).
4. Muthuri, S. G. *Lancet Resp. Med.* [http://dx.doi.org/10.1016/S2213-2600\(14\)70041-4](http://dx.doi.org/10.1016/S2213-2600(14)70041-4) (2014).



People walk along a road damaged in the 1 April Chilean earthquake.

#### SEISMOLOGY

# Chile quake defies expectations

*Smaller-than-expected tremor has scientists scrambling to redefine rules for areas of extreme seismic stress.*

BY ALEXANDRA WITZE

Monika Sobiesiak wasn’t expecting the morning of 2 April to start with such an adrenaline jolt. But as she scrolled through a list of earthquakes on her mobile phone, she saw that overnight a series of quakes had rocked the coast of northern Chile — almost exactly where she had installed a seismometer network a few years earlier. “I saw the 8.2,” says the geophysicist, who works at the University of Kiel in Germany, “and I rushed to get to my desk.”

That 1 April quake, which struck offshore near the village of Pisagua, was the largest in Chile since a magnitude-8.8 quake hit farther south in 2010. Although the Pisagua quake was not as big and not particularly damaging, it will still go down in the annals of seismology — as an intensively studied earthquake that upends some assumptions about how and when big quakes happen.

In one sense, seismologists knew it was coming. Northern Chile, near the border with Peru, was the only stretch of the country’s coastline that had not broken in a large earthquake in the past century (see ‘Under pressure’). In 2006, expecting it to go, a German–French–Chilean collaboration blanketed the

region with seismometers, tiltmeters and other ground-measuring instruments, creating the Integrated Plate boundary Observatory Chile (IPOC). It captured the Pisagua quake in action, as did Sobiesiak’s network.

But the earthquake was not the ‘Big One’ that seismologists had expected. Only a monstrous earthquake, of around magnitude 9, would have relieved all the geological stress built up in the region. More quakes, on the order of

“A lot of energy remains to be released in north Chile.”

magnitude 8, are still possible, but when they might strike is a mystery. More broadly, the Pisagua event has seismologists rethinking some basic ideas about the risk of earthquakes in similar geological settings elsewhere — places with deep-diving crustal plates, such as Japan and Indonesia.

Over time, earthquakes rupture particular portions of a long fault zone; the unbroken portions are ‘seismic gaps’ considered ripe for future quakes. Officials in these areas are often told to prepare for the worst-case scenario — the biggest possible earthquake in a given seismic gap. But the Pisagua quake shows that this does not always happen, says Susan Beck, a seismologist at the University of Arizona in

IVAN ALVARADO/REUTERS/CORBIS

Tucson. Instead, it underscores that seismic gaps can rupture in all sorts of ways, from lots of smaller quakes to just a few big ones.

Chile is an ideal laboratory in which to study such questions because it lies on the margin of a subduction zone, where the Nazca tectonic plate dives — or subducts — beneath the South American plate. Geological stress builds up and then is released in the occasional massive jolt. Chile is home to the largest earthquake ever recorded — one of magnitude 9.5 in 1960 — and accounts for more than one-quarter of the planet's total seismic-energy release.

Pisagua had not seen a major earthquake since 1877, when a tremor of around magnitude 9 ripped through the area. Seismic activity began to pick up last August, when a swarm of small earthquakes struck the area. Another set followed over the new year, and a third cluster occurred in March. These three swarms seem to have prepared the subduction zone to rupture in the big 1 April quake, says Onno Oncken, a geophysicist at the German Research Centre for Geosciences in Potsdam.

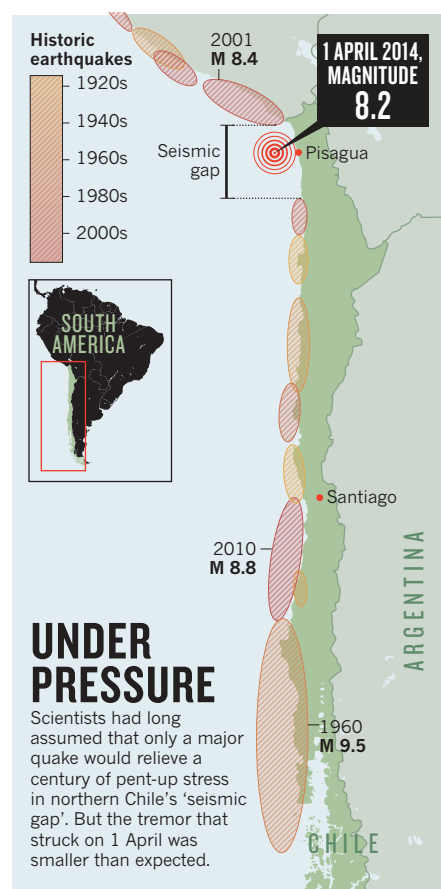
Until recently, researchers had thought that the next large earthquake in northern Chile would break the entire interface between the Nazca and South American plates, says Jean-Pierre Vilotte, a seismologist at the Paris Institute of Earth Physics. The Pisagua quake and a magnitude-7.6 aftershock two days later, “are a clear counterexample of this

simplistic classification”, he says. Together they ruptured just a small portion of the entire region at risk.

Intriguingly, the part of the subduction zone that broke was not the part that had built up the most stress, according to a ground-motion study of northern Chile's seismic gap by Marianne Métois, a geophysicist at the National Institute of Geophysics and Volcanology in Rome, and her colleagues (M. Métois *et al.* *Geophys. J. Int.* **194**, 1283–1294; 2013). For some reason, the Pisagua quake released stress in areas that were not the most wound up. “A lot of energy remains to be released in north Chile,” Métois says.

When the next one comes, seismologists plan to be ready. IPOC has added instruments to capture aftershocks from the 1 April quake and whatever might happen next. Dozens of new seismometers and global-positioning stations have been deployed by teams from Chile, Germany and France, says Sergio Barrientos, director of the National Seismological Centre at the University of Chile in Santiago.

For Sobiesiak, every little bit of data helps. Before the 1 April quake, she thought that the northern Chile seismic gap would rupture either to the north or to the south of the Pisagua area, but not right through it. Now she has some fresh thinking to do. “Each of these efforts really does bring us a step forward,” she says. ■



SOURCE: R. PRITCHARD & R. ALLMENDINGER, CORNELL UNIV.

## PHYSICS

# Quantum communications leap out of the lab

*China begins work on super-secure network as ‘real-world’ trial successfully sends quantum keys and data.*

BY JANE QIU

Cybersecurity is a step closer to the dream of sending data securely over long distances using quantum physics — spurred by two developments.

This week, China will start installing the world's longest quantum-communications network, which includes a 2,000-kilometre link between Beijing and Shanghai. And a study jointly announced this week by the companies Toshiba, BT and ADVA, with the UK National Physical Laboratory in Teddington, reports “encouraging” results from a network field trial, suggesting that quantum communications could be feasible on existing fibre-optic infrastructure.

Conventional data-encryption systems rely on the exchange of a secret ‘key’ — in binary 0s and 1s — to encrypt and decrypt information. But the security of such a communication channel can be undermined if a hacker ‘eavesdrops’ on this key during transmission. Quantum communications use a technology called quantum key distribution (QKD), which harnesses the subatomic properties of photons to “remove this weakest link of the current system”, says Grégoire Ribordy, co-founder and chief executive of ID Quantique, a quantum-cryptography company in Geneva, Switzerland.

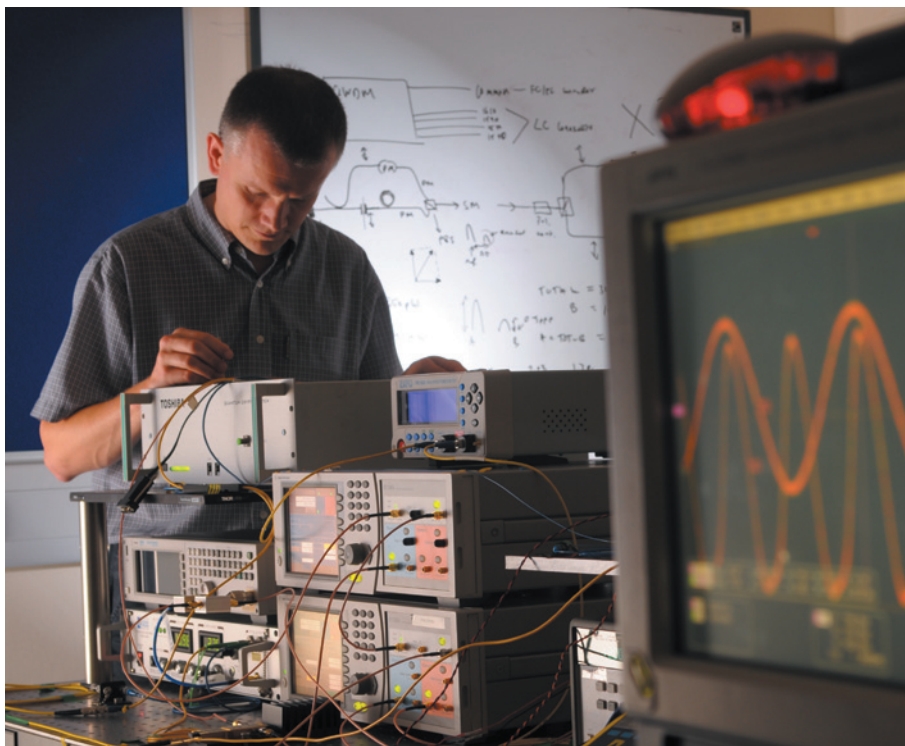
The method allows a user to send a pulse of photons that are placed in specific quantum states that characterize the cryptographic key. If anyone tries to intercept the key, the

act of eavesdropping intrinsically alters its quantum state — alerting users to a security breach. Both the US\$100-million Chinese initiative and the system tested in the latest study use QKD.

The Chinese network “will not only provide the highest level of protection for government and financial data, but provide a test bed for quantum theories and new technologies”, says Jian-Wei Pan, a quantum physicist at the University of Science and Technology of China in Hefei, who is leading the Chinese project.

Pan hopes to test such ideas using the network, along with a quantum satellite that his team plans to launch next year (see *Nature* **492**, 22–25; 2012). Together, he says, the technologies could perform further tests of fundamental





Engineer Andrew Sharpe works on Toshiba's quantum key distribution system in Cambridge, UK.

► quantum theories over large scales (around 2,000 kilometres), such as quantum non-locality, in which changing the quantum state of one particle can influence the state of another even if they are far apart, says Pan.

Sending single photons over long distances is one of the greatest problems in QKD because they tend to get absorbed by optical fibres, making the keys tricky to detect on the receiver's end.

This is “a big challenge for conventional detectors”, says Hoi-Kwong Lo, a quantum physicist at the University of Toronto in Canada. But technological breakthroughs in recent years have significantly reduced the noise level of detectors while increasing their efficiency in detecting photons from just 15% to 50%.

Vast improvements have also been made in the rate at which detectors can ‘count’ photon pulses — crucial in determining the rate at which quantum keys can be sent, and thus the speed of the network. Counting rates have been raised 1,000-fold, to about 2 gigahertz, says Lo.

The breakthroughs are pushing the distance

over which quantum signals can be sent. Trials using ‘dark fibres’ — optical fibres laid down by telecommunications companies but lying unused — have sent quantum signals up to 100 kilometres, says Don Hayford, a researcher at Battelle, a technology-development company headquartered in Columbus, Ohio.

To go farther than that, quantum signals must be relayed at ‘node points’ — the quantum networks between Beijing and Shanghai, for instance, will require 32 nodes. To transmit photons over longer distances without the use of nodes would require a satellite.

China is not alone in its quantum-communication efforts. A team led by Hayford, together with ID Quantique, has started installing a 650-kilometre link between Battelle's headquarters and its offices in Washington DC. The partnership is also planning a network linking major US cities, which could exceed 10,000 kilometres, says Hayford, although it has yet to secure funding for that.

The Chinese and US networks will both use dark fibres to send quantum keys. But these fibres “are not always available and can be

prohibitively expensive”, says Andrew Shields, a quantum physicist at Toshiba Research Europe in Cambridge, UK. One way to sidestep the problem is to piggyback the photon streams onto the ‘lit’ fibres that transmit conventional telecommunications data. However, those conventional data streams are usually about a million times stronger than quantum streams, so the quantum data tend to be drowned out.

In the results announced this week, Shields and his colleagues were successful in achieving the stable and secure transmission of QKDs along a live lit fibre between two stations of the UK telecommunications company BT, 26 kilometres apart. The quantum keys were sent over several weeks at a high rate alongside four channels of strong conventional data on the same fibre.

The research builds on previous work in which Shields and his team developed a technique to detect quantum signals sent alongside noisy data in a 90-kilometre fibre, but in controlled laboratory conditions (K. A. Patel *et al.* *Phys. Rev. X* 2, 041010; 2012).

“Implementing QKD in the ‘real world’ is much more challenging than in the controlled environment of the lab, due to environmental fluctuations and greater loss in the fibre,” says Shields.

The quantum keys in the latest study were sent alongside conventional data travelling at

**The breakthroughs are pushing the distance over which quantum signals can be sent.**

40 gigabits per second. “As far as I am aware, this is the highest bandwidth of data that has been multiplexed with QKD to date,” add Shields.

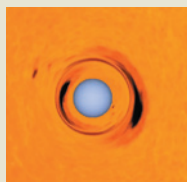
He calculates that it would be possible to send QKD signals alongside 40 conventional data channels. Optical fibres usually carry between 40 and 160 telecommunications channels, meaning that quantum communication could be carried out with existing infrastructure.

“I find it an impressive piece of work that demonstrates the multiplexing of strong classical signals with quantum signals in the same fibre for the first time” in a field trial, says Lo. Removing the need for dark fibres, he says, is an important step in showing that QKD has the potential to be used in “real life”. ■

TOSHIBA RESEARCH EUROPE

  
**MORE ONLINE**

#### VIDEO OF THE WEEK



Planet hunter sees gravitational self-lensing in double star  
[go.nature.com/l3mlxy](http://go.nature.com/l3mlxy)

#### MORE NEWS

- An Earth-like planet in the habitable zone [go.nature.com/tcxyzc](http://go.nature.com/tcxyzc)
- Insect's ‘female penis’ extracts nourishment with semen [go.nature.com/cjuckd](http://go.nature.com/cjuckd)
- Archaic human epigenomes get mapped [go.nature.com/hnzsye](http://go.nature.com/hnzsye)

#### NATURE PODCAST



The worth of the Y chromosome; an unusual amateur scientist; and how to help corals breed [nature.com/nature/podcast](http://nature.com/nature/podcast)



## BIOTECHNOLOGY

# RNA interference rebooted

*Gene-silencing technique yields promising treatments for liver-linked disorders.*

BY ERIKA CHECK HAYDEN

After a rocky start, RNA interference (RNAi), a gene-silencing technique that won the 2006 Nobel Prize in Physiology or Medicine, is gaining momentum.

In November, Alnylam Pharmaceuticals in Cambridge, Massachusetts, announced its first major success: evidence that an RNAi-based treatment drastically reduced levels of a toxic liver protein in people with a rare neurodegenerative disease. That opened up the floodgates: biotechnology firms have already raised nearly US\$1 billion this year for RNAi-based drug development, and roughly 150 therapies are being tested in patients. "It's no longer a question of if RNA therapeutics will become a reality" but when, says John Maraganore, Alnylam's chief executive.

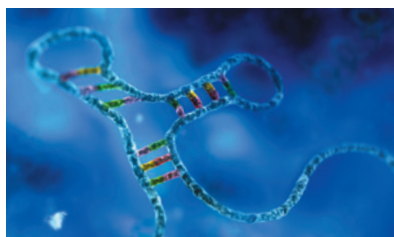
Dirk Haussecker, a biotechnology analyst in Baden-Baden, Germany, says that Alnylam and other companies pursuing RNAi therapies have overcome two significant challenges. First, they have solved the problem of delivering therapeutic snippets of RNA to the right cells at the right time. Second, they have narrowed their focus largely to diseases of the liver, exploiting its cells' natural function: absorbing large molecules from the blood. "For a long time, people doubted that this technology could yield a drug that is commercially viable, but that doubt has been removed over the past two years by clinical results," says Haussecker, author of the blog RNAi Therapeutics (see [go.nature.com/gtekcc](http://go.nature.com/gtekcc)).

Still, sceptics argue that the technology is not ready for prime time: with only limited results from clinical trials of RNAi therapies, it is not yet certain that the approach is free of significant side effects. Naysayers also note that several major pharmaceutical companies have abandoned RNAi research since 2010. Most recently, Novartis in Basel, Switzerland, said on 14 April that it is substantially scaling back its RNAi research programme, citing "ongoing challenges with formulation and delivery" of RNAi therapies and the prospect that they may work for only a narrow range of conditions.

RNAi was discovered in 1998 by two US geneticists: Andrew Fire, now at Stanford University, California, and Craig Mello of the University of Massachusetts Cancer Center in Worcester. The technique involves infusing a cell with small pieces of custom-made RNA. These bind to and disrupt the native RNA that translates DNA into proteins. Scientists have long hoped that RNAi could block mutant genes that encode faulty proteins that cause disease.

## UPS AND DOWNS

Large drug companies have abandoned studies of RNA interference (RNAi), but smaller biotech firms are pushing ahead.



**OCTOBER 2006** US geneticists Andrew Fire and Craig Mello win Nobel prize for RNAi (RNA molecule pictured).

**OCTOBER 2006** Merck pays US\$1.1 billion for California-based RNAi pioneer Sirna Therapeutics.

**MARCH 2009** OPKO Health of Miami, Florida, terminates phase III trial of RNAi therapy for macular degeneration.

**NOVEMBER 2010** Swiss firm Roche ends RNAi research after 3 years and \$500 million.

**NOVEMBER 2013** Alnylam reports successful phase II trial of RNAi-based drug for liver-linked disease.

**JANUARY 2014** Alnylam buys Sirna from Merck for \$175 million.

**APRIL 2014** Novartis shuts most RNAi research.

But getting delicate, small RNAs into cells has proved difficult. And the first major RNAi clinical trial — which aimed to treat macular degeneration — was halted in 2009 after reports that the therapy might actually cause blindness. The small RNAs used in the trial, delivered with little modification, triggered cell death through an immune pathway that can be activated by any RNA more than 21 bases long.

RNAi-based therapies fell out of favour after the failed trial, but smaller firms kept working on delivery systems. One such system, developed by Tekmira in Burnaby, Canada, is based on lipid nanoparticles. Patisiran, the drug made by Alnylam to treat a rare disease known as transthyretin-mediated amyloidosis, is an

intravenous infusion of small RNAs packaged in these nanoparticles, which resemble a suspension of very small oil droplets. The drug is intended to reduce the production of a mutant liver protein that circulates in the blood and accumulates in other tissues, causing nerve damage and, in many cases, death.

In a 29-patient trial, Alnylam showed that patisiran could reduce blood levels of the protein, transthyretin, by as much as 96%. The company has begun an expanded trial that it plans to conclude in 2017. It is also testing other RNAi-based drugs using its own delivery vehicle, a sugar molecule that is attached to a small RNA and targets a receptor on liver cells.

Another firm, Arrowhead Research of Pasadena, California, has devised a third major delivery system — a customizable polymer that deposits RNAs in target cells. It is now being tested in clinical trials. "We've spent a lot of time in the wilderness trying to solve the delivery problem," says Christopher Anzalone, Arrowhead's chief executive.

Victor Kotelianski, a former Alnylam executive who directs the RNA centre at the Skolkovo Institute of Science and Technology in Russia, says that caution is warranted. Larger trials of patisiran might uncover side effects that have not yet surfaced, if the RNAi approach is not as precise at hitting its targets as early results suggest. The company has also not yet proved that removing faulty transthyretin proteins from the blood actually resolves patients' symptoms, he says.

There are still enough lingering questions about RNAi's usefulness to give large firms pause. "With a technology like RNAi, you don't necessarily know what its applications in the clinic will be, and those applications might not be compatible or consistent with the goals of the big pharmaceutical companies," Haussecker says. Before Novartis's decision to abandon most of its RNAi drug development, Merck of New Jersey disclosed on 12 January that it had sold its small-RNA drug-discovery programme to Alnylam for \$175 million — much less than the \$1.1 billion that Merck paid to kick-start its RNAi research in 2006 (see 'Ups and downs').

Yet Maraganore remains bullish, drawing a parallel between RNAi therapies and monoclonal antibodies — large, complex proteins that bind to specific cell receptors, which were brought to market by focused biotech companies, not big drug firms. "Large companies don't do a good job with technology," Maraganore says. "It's the same story as with antibodies — we're seeing that movie play out all over again." ■





# DESIGNER REEFS

Biologists are directing the evolution of corals to prepare them to fight climate change.

BY AMANDA MASCARELLI

Off the coast of American Samoa, the tropical sun beats down on a shallow tidal lagoon, heating the water to a sizzling 35 °C for a few hours each day. Such temperatures would kill off most coral reefs, and yet the Samoan lagoon hosts courtyards of antler-like branching corals and mound corals the size of refrigerators. “The fact that they’re there means they’ve adapted to survive,” says Steve Palumbi, a marine biologist at Stanford University in California. “The real question is: how did they do that and can all corals do that?”

Palumbi is just starting to understand how these Samoan corals thrive in such extreme conditions. And he thinks he might be able to harness that ability to create a reef of hardy coral with a chance of surviving the hot seas that are expected to result from climate change. Starting in August, he and his team are going to try to plant “the smartest future reef we can imagine”.

Palumbi is part of a small group of coral researchers around the world tackling such issues to throw threatened reefs a lifeline. Their ultimate intent is to launch a programme of ‘human-assisted evolution’, creating resistant

**Reefs thrive in the hot waters of American Samoa that would kill other corals.**

FLORIS VAN BREUGEL/NATUREPL.COM



corals in controlled nurseries and planting them in areas that have been — or will be — hard-hit by changing conditions. “It’s a brave new world of working with corals in this way,” says Ruth Gates, a marine biologist at the University of Hawaii at Manoa who, along with coral geneticist Madeleine van Oppen at the Australian Institute of Marine Science in Townsville, is helping to pioneer the field.

The work is not without controversy. Although no one is yet attempting to create genetically modified corals, some researchers are concerned that human-assisted evolution goes too far down the slippery slope of altering natural systems. “If you’re basically farming a reef, you’ve taken a natural habitat and you’ve converted it,” says Steve Vollmer, a coral geneticist at Northeastern University’s Marine Science Center in Nahant, Massachusetts, who feels that more needs to be known before embarking on such programmes. “It’s like going to the Midwest and taking grasslands and making it into soy. There are huge implications to doing this.”

### IN HOT WATER

Coral reefs have been besieged in recent decades by everything from warming waters to ocean acidification, disease, overfishing and pollution. According to *Status of Coral Reefs of the World: 2008*, a synthesis report<sup>1</sup> by hundreds of scientists and environmental managers, 19% of the world’s coral reefs have been lost since 1950 and another 35% are threatened or in critical condition. Some areas have suffered disproportionately: the Caribbean, for example, has lost 80% of its reefs since the 1970s (ref. 2). By the end of this century, researchers expect ocean waters to drop from a pH of 8.1 to 7.9 or lower, and to warm by at least 2 °C, averaged across the globe. “It’s kind of like if you pull the plug on the bathtub and the water is rushing out — that’s the state of corals,” says Palumbi.

Reef-restoration projects have been focusing on the Caribbean and other hard-hit spots for more than 20 years. In these programmes, small samples are taken from local reefs and grown in controlled coral nurseries. After a few months, fragments the size of a hand or larger can be ‘outplanted’ to a reef using underwater cement, where the coral will continue to grow. Such projects have shown that transplantation and reef restoration can be done on a small scale. But transplanted corals grow more slowly and have higher mortality rates than normal<sup>3</sup>. “Coral restoration has always been highly expensive and slow and inefficient,” says Palumbi. “Figuring out how to do this in a smarter way is our goal.”

That smarter way takes advantage of the surprising resilience and resourcefulness of some corals and the symbiotic algae that live inside them. “Sometimes we find reefs that are doing very, very well in places that you would least expect to find them,” says Gates — such as a reef off Taiwan that lies below the waste-water outfall pipe of a nuclear power plant and experiences temperature fluctuations of between 6 °C and 8 °C per day. “By all of our understanding, we would expect those corals to all be dead. But they’re not, they’re flourishing.”

Waters with a reduced pH are expected to dissolve coral skeletons — but in Palau in the western Pacific Ocean, researchers have found<sup>4</sup> reefs that are bigger and more diverse in relatively acidic waters than the Pacific average. Another study<sup>5</sup> found that dire predictions about the frequency of future coral-bleaching events — mass die-offs when stressed corals lose their symbiotic algae — are reduced by 20–80% if the models take into account corals’ ability to adapt after previous bleaching events. That delays predicted mass reef deaths by about a decade.

So far, researchers have only a handful of hints as to what makes some corals resilient. In a study<sup>6</sup> published in 2013, Palumbi and his colleagues, including Daniel Barshis, a marine biologist at Stanford, compared two populations of the reef-building coral *Acropora hyacinthus* at their field site off Ofu Island in American Samoa. One population lives in the toasty pool where temperatures reach 35 °C during summer low tides and fluctuate by up to 6 °C daily; the other, less isolated by tides, has to deal with temperatures of only about 29 °C. The team placed samples in controlled tanks and shocked them with temperatures of nearly 3 °C above normal for

four days. All of the corals bleached by the end of the fourth day. But those from the hotter pool survived for longer and had higher expression of 60 genes, including well-known thermal-tolerance genes such as those that make heat-shock proteins and antioxidant enzymes.

Palumbi and Barshis think that genetic fitness and acclimatization both play important roles in boosting tolerance. Their analyses suggest that corals can ‘toughen up’ over the course of their lifetimes in response to environmental conditions. Those in the hot pool are physiologically primed to tolerate additional heat stress, “like an athlete who’s been training every day since a very early age”, says Barshis.

## “WE WOULD EXPECT THESE CORALS TO ALL BE DEAD. BUT THEY’RE NOT, THEY’RE FLOURISHING.”

A promising twist is that the more heat-tolerant species seem also to be more transplant-friendly. After experimentally planting some 400 samples from the two reef areas back into the two pools, Palumbi and his team found that the corals from the hotter pools transplanted more efficiently and grew faster than those from the cooler pools.

This August, Palumbi and his colleagues plan to begin an experimental restoration project on Sili Reef off Ofu Island. To select the best corals, the researchers will rely on their extensive data for the area, including growth measurements and transcriptomes — blueprints of the part of the genome that is actively transcribed into proteins. They also plan to use data from a portable stress test for corals that Palumbi is developing — “like a human treadmill test for cardiac function”, he says. He and his team have built tanks out of 7.5-litre cooler boxes rigged with lights, heaters and chillers that can dose corals with a controlled bout of high physiological stress. By monitoring bleaching and chlorophyll content, they should be able to predict how corals might respond to potential bleaching conditions.

Using all this information as a guide, they will handpick the hardest, fastest-growing and most heat-resistant corals for their smart reef. At the same time, they will build a second reef from corals selected at random. They will then monitor reef survival over several years. “The question is: can we do better if we have a lot of information about the individual corals?” says Palumbi. “Honestly, I don’t know the answer.”

### LEGACY OF SURVIVAL

Others have found encouraging evidence that stress resistance gained through acclimatization can be passed on to offspring. Unpublished work by Gates, led by the University of Hawaii’s Hollie Putnam, shows that adult cauliflower corals (*Pocillopora damicornis*) exposed to stress during brooding produce larvae with increased resilience to heat and ocean acidification. The team hypothesizes that this transgenerational protection is caused by epigenetic changes: the modification of molecular tags on the genome that affect gene expression.

Gates and van Oppen are aiming to look specifically at areas that have already survived massive bleaching events, such as Moorea in French Polynesia, the central Great Barrier Reef in Australia, and the Seychelles, where 97% of corals in the inner islands died following the 1997–98 El Niño oceanic warming event. (A nursery has already been created from the Seychelles corals that survived, and fragments grown from them have been planted onto reefs to aid their recovery.) Gates and van Oppen aim to cross-breed corals that have survived such stressful bleaching, and to track the resilience of the offspring.

Their ideas won Gates and van Oppen the 2013 Paul G. Allen Ocean

➔ NATURE.COM

To hear more about how researchers are helping corals, visit: [go.nature.com/ov46wo](http://go.nature.com/ov46wo)





Madeleine van Oppen collects corals for study and selective breeding.

Challenge prize of US\$10,000, along with an invitation to apply for multimillion-dollar funding. Depending on how much of that funding comes through, they also aim to use heat and acidity to stress corals before they breed, to see if and how tolerance gets passed down the generations. Beginning in May, van Oppen and her team will start collecting adults of the branching coral *Pocillopora acuta* from the Great Barrier Reef, and will grow them in the Australian Institute of Marine Science's massive National Sea Simulator, an aquarium facility that provides controlled tanks to replicate open-ocean conditions.

Ultimately, Gates and van Oppen hope to create a 'seed bank' of gametes and fertilized embryos from extreme settings in which corals persist despite the odds — including the shallow reefs skirting Coconut Island, Hawaii, where both temperature and pH fluctuate drastically, reaching upper limits similar to those expected in the open ocean by 2050. The seed bank would add to efforts spearheaded by the US Smithsonian Institution, in collaboration with Hawaiian and Australian bodies, which are already banking coral sperm and embryonic cells.

A final, important piece of the puzzle is the corals' symbiotic algae: these are shorter-lived and faster-evolving than their hosts, and research has shown that they can pass along thermal tolerance. One study<sup>7</sup>, for example, found that juvenile corals inoculated with strains of algae collected from a warm reef known for heat resistance grew well when exposed to temperatures up to 32 °C, whereas samples of the same coral inoculated with algae from a cooler reef suffered bleaching and tissue death.

Andrew Baker, a marine biologist at the University of Miami in Florida, and his colleagues discovered<sup>8</sup> that symbionts from a lineage called clade D tend to become more prevalent in some corals when they are heat stressed, suggesting that the algae are better able than other strains to survive such conditions, and that they help their hosts to survive too. Since then, studies have shown<sup>9</sup> that clade D symbionts, in particular types D1 and D1a, are prevalent in a wide variety of corals that have survived extreme bleaching events. Putnam, Gates and their colleagues

have found<sup>10</sup> that a different strain, C15, seems to be dominant in heat-resistant corals near Moorea.

Researchers such as Baker are starting to think about the possibility of intentionally seeding coral reefs with hardier strains of algae to help them to resist the perils of climate change. But it is still unclear whether it will be possible to manipulate symbiont populations effectively in the wild, where environmental conditions might cause the corals to favour one type of algae over another.

Workers at existing coral nurseries and farms have been sending samples of coral and symbionts to researchers for genetic sequencing, while keeping tabs on which organisms fare well in heat shocks or disease outbreaks. Researchers have banked hundreds of genotyped strains from a handful of coral species, including the critically endangered staghorn coral (*Acropora cervicornis*) from locations in the Caribbean, says Les Kaufman, a marine biologist at Boston University in Massachusetts.

## HELP OR HARM

The days of trying to build reefs with designer-made corals are still in the future. But as the research heads in that direction, some are wary that such tinkering might do more harm than good.

Selecting for traits such as resistance to heat or acidification might lead to a genetic bottleneck, for example. "Selective-breeding programmes may effectively reduce the capacity of corals to adapt to future changes in environmental conditions by narrowing genetic variation," says David Miller, a coral biologist at James Cook University in Townsville. And that is if selective breeding in corals even works. It is too soon even to tell whether acid and heat resistance are strongly heritable, he says.

Miller and others point out that cross-breeding to enhance specific traits in crops and dogs, for example, often comes at the expense of other traits. "There's often a 'trade-off' effect, so that, for example, more-stress-tolerant individuals are likely to grow more slowly," says Miller. Selecting for resilience against heat and acidity could hypothetically lead to higher susceptibility to disease, for instance.

Manuel Aranda, an evolutionary molecular biologist at the Red Sea Research Center at King Abdullah University of Science and Technology in Thuwal, Saudi Arabia, agrees that breeding might come at a cost. But he says that the serious decline in reef health warrants exploring all available options. "If you think about losing an entire ecosystem, you want to start somewhere."

Until recently, says Baker, the goal for coral-reef management was simply to create marine reserves and reduce the pressures of pollution and fishing, hoping that that would leave reefs strong enough to deal with climate change. "The pendulum has sort of shifted as people have realized just how dire the situation is," says Baker. "We need to do more than that. We need to take action."

Some 500 million people depend in some way on coral reefs for food and income, and the livelihoods of another 30 million are entirely dependent on reefs<sup>1</sup>. For Gates, statistics like those, combined with the facts of climate change, make the pursuit of assisted evolution necessary and urgent. "We don't have a lot of time," she says. ■

**Amanda Mascarelli** is a freelance reporter based in Denver, Colorado.

1. Wilkinson, C. (ed.) *Status of Coral Reefs of the World: 2008* (Global Coral Reef Monitoring Network and Reef and Rainforest Research Centre, 2008); available at [go.nature.com/6ffbar](http://go.nature.com/6ffbar).
2. Gardner, T. A., Côté, I. M., Gill, J. A., Grant, A. & Watkinson, A. R. *Science* **301**, 958–960 (2003).
3. Edwards, A. J. & Clark, S. *Mar. Pollut. Bull.* **37**, 474–487 (1999).
4. Shamberger, K. E. F. et al. *Geophys. Res. Lett.* **41**, 499–504 (2014).
5. Logan, C. A., Dunne, J. P., Eakin, C. M. & Donner, S. D. *Glob. Chang. Biol.* **20**, 125–139 (2014).
6. Barshis, D. J. et al. *Proc. Natl Acad. Sci. USA* **110**, 1387–1392 (2013).
7. Howells, E. J. et al. *Nature Clim. Change* **2**, 116–120 (2012).
8. Baker, A. C., Starger, C. J., McClanahan, T. R. & Glynn, P. W. *Nature* **430**, 741 (2004).
9. LaJeunesse, T. C., Smith, R. T., Finney, J. & Oxenford, H. *Proc. R. Soc. B* **276**, 4139–4148 (2009).
10. Putnam, H. M., Stat, M., Pochon, X. & Gates, R. D. *Proc. R. Soc. B* **279**, 4352–4361 (2012).

# BIOFUELS HEAT UP

*A new generation of industrial plants can make liquid fuels from almost any organic scraps — from corn stalks and wood chips to urban rubbish.*

BY KIM KRIEGER

**B**y the end of 2015, all British Airways flights out of London City Airport will be fuelled by rubbish — the paper, food scraps, garden clippings and other organic detritus discarded by the city's residents.

But before it goes into planes, the rubbish will be processed at GreenSky London: a biofuels plant under construction on the eastern side of the city. Each year, the facility will take in some 500,000 tonnes of the city's waste and will transform the organic component into 60,000 tonnes of jet fuel, a similar quantity of diesel fuel combined with petrol-like naphtha, and 40 megawatts of power.

This level of output would hardly be noticed at conventional petroleum refineries, which typically generate a similar volume of product within a week. But "gathering enough biomass to run a petroleum-scale refinery is almost unthinkable", says Nathanael Greene, director of renewable-energy policy at the Natural Resources Defense Council in New York City. GreenSky London is typical of a trend for second-generation biofuel reactors that are not only omnivorous — they can be fed with corn stalks, wood chips and other forms of agricultural waste, as well as urban rubbish — but also small. The hope is that they will slash transportation costs by bringing the reactors to the biomass instead of vice versa.

Proponents argue that novel catalytic techniques and compact designs will make second-generation biofuel plants not just environmentally friendly, but also profitable enough to compete with petroleum-based fuels without subsidies. Questions remain about how realistic this hope is. But at least some customers are giving the plants a try; commercial units have begun to spring up from Finland to Mississippi to Alaska.

If these second-generation plants do succeed, says Greene, they will offer at least one crucial advantage over their predecessors: a low-carbon way to create fuel that suits existing vehicles.

Limited compatibility in this regard is a key problem dogging the first generation of biofuels plants, which rely on technology developed over millennia to make beer, wine and spirits. These facilities grind up edible products such as corn or sugarcane, add water and yeast, and allow fermentation to take its natural course. The result is a copious supply of ethyl alcohol, which makes an excellent fuel and can be mixed with petrol.

But there are serious drawbacks to making fuel from food in a world with a growing population and limited arable land. So, for more than a decade, the biofuels industry has been working on economical ways to use cornstalks, wood chips and other by-products that currently go to waste. This has posed a challenge for the fermentation approach, because these materials contain tough, long-chain molecules such as cellulose and lignin that yeast cannot easily digest. Over the past five or ten years,

advances involving pretreatment with acids and enzymes have partially overcome that barrier, and commercial plants designed to produce cellulosic ethanol are now under construction in Iowa and Kansas (see *Nature* 507, 152–153; 2014).

## HITTING THE WALL

But even these facilities will not be able to overcome the biggest restriction on the fermentation approach: the 'blend wall'. This is the maximum amount of ethanol that can be mixed into petrol without causing corrosion in fuel lines and car engines. For current models, the blend wall is about 10–15% — and first-generation fermentation plants already produce more than enough ethanol to meet this demand. Indeed, several US ethanol refineries built in the past decade already stand idle, victims of drought-induced price hikes and market saturation.

That reality, paired with nine years of historically high oil prices — the price per barrel currently stands at around US\$100 — has spurred vigorous research into thermochemical reactors, which convert biomass directly into fuels other than ethanol using heat and catalysts.

The most common thermochemical approach is gasification, in which carbon-rich material such as coal, wood chips or municipal waste is heated to produce synthesis gas or 'syngas' — a mixture of mainly hydrogen and carbon monoxide, with traces of carbon dioxide and other gases (see 'Fuel flow'). At the GreenSky London facility, one or more proprietary gasifier units built by Solena Fuels, a renewable energy firm in Washington DC, will accomplish this step by vaporizing the waste with jets of ionized plasma that heat the material to some 3,500 °C. Such torches are more energy intensive than other methods of gasification, in which the biomass is heated from below on beds of sand or other material. GreenSky London elected to use them because the contents of municipal waste can vary markedly, and by adjusting the temperature of the torches, the composition of the syngas can be kept consistent.

Consistency is important for optimizing the second step of the process, in which the syngas is sent into a chemical reactor — made, in the case of GreenSky, by Velocys of Plain City, Ohio. There, it undergoes the Fischer–Tropsch reaction, which fuses hydrogen and carbon monoxide into long-chain hydrocarbons. Velocys has made its system unusually compact by reducing the size of the cobalt-containing catalyst particles to the nanometre scale and arranging them along a series of microchannels, which direct the flow of the syngas and produce reaction surfaces with an effective area of a much larger device.

The Fischer–Tropsch units have also been made

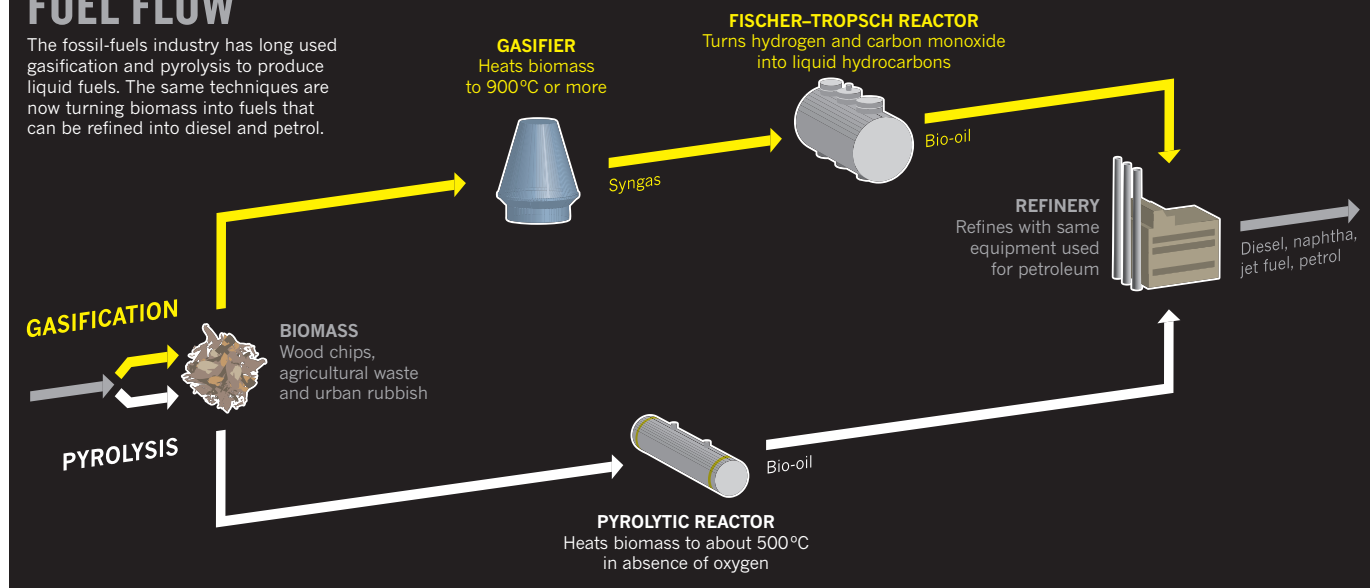
## ➔ NATURE.COM

For more on small, local sources of power, see:

[go.nature.com/nubllr](http://go.nature.com/nubllr)

## FUEL FLOW

The fossil-fuels industry has long used gasification and pyrolysis to produce liquid fuels. The same techniques are now turning biomass into fuels that can be refined into diesel and petrol.



as modular as possible, so that chunks can be assembled at a factory and then be plugged together on site. “You need to achieve economy not through size, but in the way you build them,” says Neville Hargreaves, business-development manager at Velocys.

Another compact system is the BioMax gasifier developed by the Community Power Corporation in Englewood, Colorado. The company says that this device is modular and small enough that four can fit into a standard shipping container, and can run on almost any kind of shredded biomass, from food scraps to cardboard to wood chips. The resulting syngas can then be used in place of natural gas for heating, cooling or electricity generation. A typical unit generates about 150 kilowatts, enough to power between 25 and 50 homes, run three supermarkets or keep vital hospital equipment working. And in the near future, BioMax units should be able to plug in a Fischer–Tropsch reactor and produce biodiesel as well.

In 2011, Community Power was bought by the Afognak Native Corporation, which is owned by the indigenous people of Alaska’s Afognak Island. They hope to sell the units throughout Alaska and northern Canada, where electricity and transportation fuel are expensive.

### CLEAN COMBUSTION

Among the strongest selling points of the two-step gasification approach to biofuels is the fact that almost all the syngas gets turned into hydrocarbons with no double bonds or ring structures, producing fuels that burn cleanly and completely. But that advantage has not kept researchers from exploring a single-step alternative. In the pyrolysis approach, the biomass is heated in the absence of oxygen to some 500 °C and converted into organic liquids directly. These liquids can then be refined into fuels using standard technology. Pyrolysis is relatively immature compared with gasification, says Mark Nimlos, a principal scientist at the National Renewable Energy Laboratory in Boulder, Colorado. But that can be seen as a virtue, he adds. “There is lots of potential to improve.”

Several companies are already testing the commercial viability of the technology. For example, UOP of Des Plaines, Illinois — a subsidiary of the New Jersey-based conglomerate Honeywell International — is partnering with Ensyn Technologies of Ottawa to market Ensyn’s Rapid Thermal Processing (RTP) units. The companies foresee these units being installed next to lumber mills, where each one would be capable of turning waste wood into some 76 million litres of pyrolysis oil a year. That would be enough to warm 31,000 homes if it were burned directly as heating oil; alternatively, if refined into petrol, it could fuel about 35,000 typical US automobiles.

Green Fuel Nordic, a biorefining company based in Kuopio, Finland, is planning to install at least one RTP unit in the Finnish town of Iisalmi, where it will process waste from the country’s extensive forestry industry. The company is also working with the European Commission to develop a set of quality standards for pyrolysis fuels. One component of concern is tar: a gummy residue of long-chain molecules that are hard to refine. Another is oxygen, which is abundant in biomass and reacts with pyrolysis oil to form organic acids that can seriously corrode refinery equipment. Finding better ways to deal with both of these contaminants is a major goal of pyrolysis-oil research. At present, the easiest way to remove the oxygen is to add molecular hydrogen derived from natural gas, but that would both undermine the climate-friendly appeal of pyrolysis oil and drive up the cost.

Indeed, the financial viability of any of the second-generation biofuel technologies is still an open question. Witness the saga of one of the world’s most advanced pyrolysis biorefineries: a \$225-million facility in Columbus, Mississippi. Owned by KiOR, a renewables company based in Pasadena, Texas, the facility demonstrated its technical viability by producing some 3.5 million litres of petrol and diesel fuel from wood waste in 2013 — about as much as a conventional petroleum refinery produces in a day. But KiOR, which shut down the Columbus refinery for upgrades in January, will run out of operating funds by the end of August — and it has enough money to last that long only because it secured a \$25-million loan earlier this month from billionaire venture-capitalist Vinod Khosla, whose company, Khosla Ventures, originally funded the project.

For GreenSky London, economic viability also remains an open question. But its partners — Velocys, Solena and British Airways — are hopeful. They have not disclosed the cost of the facility, but none of them see cost as the central issue. British Airways expects the operation to help it meet the carbon-emissions targets mandated by the European Union while ensuring a steady supply of jet fuel that is not subject to the price fluctuations that plague the oil market. And both Solena and Velocys hope that GreenSky London will be the first of many such facilities serving airports around the world.

Every field, forest and landfill site is a potential fuel source for these facilities, says Hargreaves. And the need for liquid fuels will never completely go away. “Fifty years in the future, we might get land transportation entirely electrified,” he says. But aircraft require a level of energy density that batteries simply cannot provide. Liquid fuel, he says, “is very difficult to substitute.” ■

**Kim Krieger** is a freelance writer in Norwalk, Connecticut.



# COMMENT

**HISTORY** Children's author Beatrix Potter was a scientist manqué **p.454**

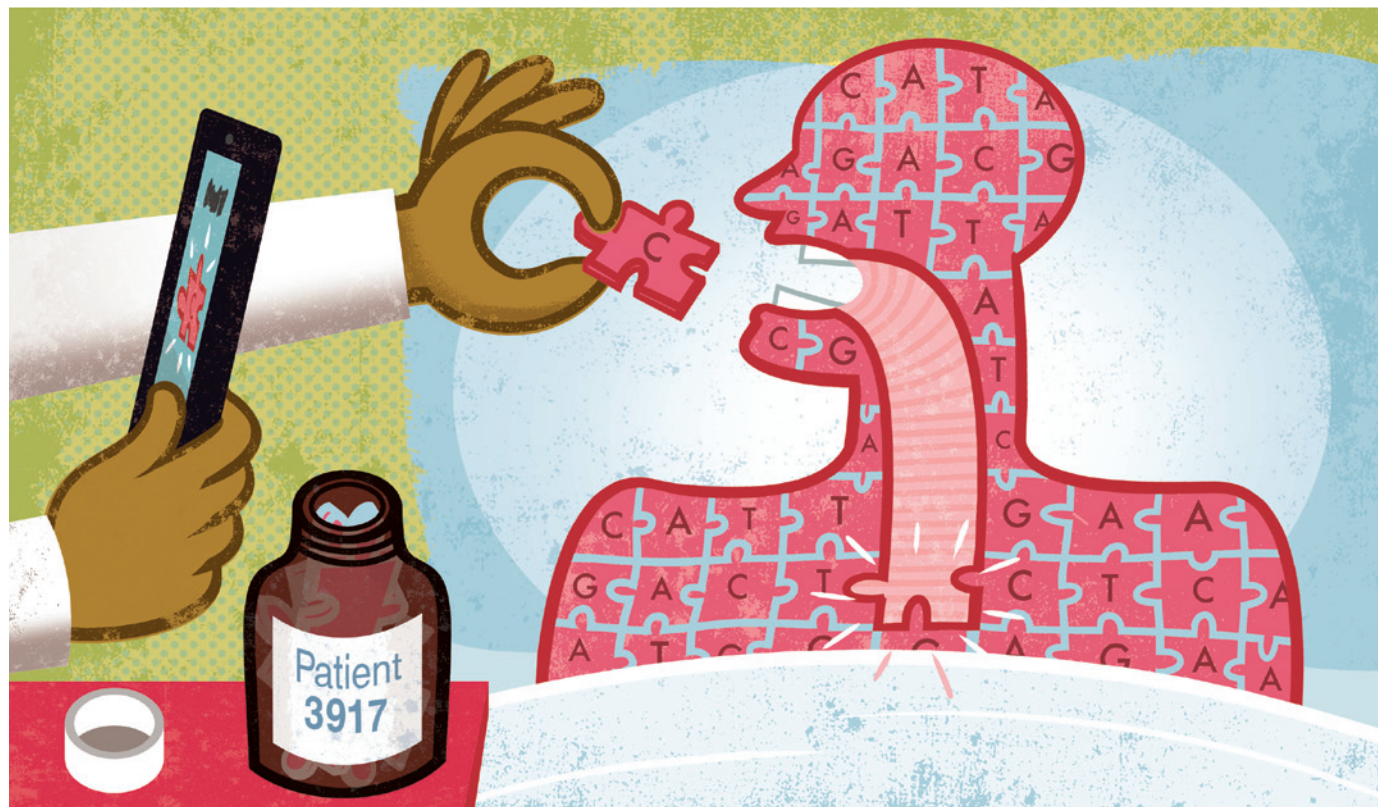


**POLICY** The rise of banks for human blood, milk and sperm **p.456**

**ATMOSPHERES** A warning about geoengineering, in three parts **p.457**

**NEUROSCIENCE** Two critiques of the US National Institute of Mental Health strategy **p.458**

ILLUSTRATIONS BY PETE ELLIS/WWW.DRAWGOOD.COM



## Gather and use genetic data in health care

Research into how genetic variants can guide successful treatments must become part of routine medical practice and records, says **Geoffrey Ginsburg**.

**M**ore and more people are getting their DNA sequenced. But the use of genetic data to inform medical decisions is lagging. More than a decade since the Human Genome Project was declared complete, fewer than 60 genetic variants are deemed worthy for use in clinical care, most for severe conditions in very young children<sup>1</sup>.

These genetic variants can guide medical decisions (see 'Genes that doctors use'). By some estimates, women with certain variants

in the *BRCA* genes have about an 80% chance of developing breast cancer, leading some who carry the mutation to opt for preventive mastectomies. Screening for faulty genes involved in iron transport can alert affected individuals to a need to alter their diets to avoid developing haemochromatosis, a toxic build-up of iron that damages the liver, heart and other organs. Mutations in the *EGFR* gene can indicate whether lung cancer will respond to expensive drugs with fewer side effects than standard chemotherapy. But five

years after *EGFR* tests were commercialized, only around 6% of appropriate US patients were being genotyped, partly because their physicians were unaware of the tests<sup>2</sup>.

Clinical trials have been used to assess whether genomic information yields practical benefits. A study<sup>3</sup> of nearly 2,000 patients with HIV showed that genetic screening for a variant called *HLA-B\*5701* could help to prevent toxic reactions to the AIDS drug abacavir — a fact that is now written into US treatment guidelines. And last year, two ►

► separate randomized trials<sup>4,5</sup> evaluated whether genotyping should guide the use of warfarin, prescribed to prevent blood clots. The studies reached different conclusions, perhaps demonstrating that such trials are not always the best way to assess the use of a genetic test.

What is more, such trials can take years and millions of dollars to complete. They are usually paid for by pharmaceutical companies in the hope of selling a specific product. But only a fraction of genetic variants are tied to commercial opportunities.

Meanwhile, numbers of intriguing genetic variants are surging. Analysis of tumour genomes has revealed some 140 genes whose mutations contribute to cancer<sup>6</sup>. Saudi Arabia, the United Kingdom and the United States have all launched projects that in total will sequence about 100,000 individuals. The clinical-sequencing market has been estimated at more than US\$2 billion.

For these investments to pay off, there needs to be an efficient way to evaluate whether genetic information leads to better health care. One solution is to gather evidence during routine clinical care. In other words, health systems should embrace research. This will require hefty investment in new workflows, informatics and data-analysis techniques.

***“The clinical-sequencing market has been estimated at more than US\$2 billion.”***

With the right infrastructure, the number of clinically useful variants could grow into the thousands, making genomics crucial to medical care<sup>7</sup>. In the United States — my focus in this article — a large, fragmented health sector fosters many independent arms that can struggle to work together, in contrast to a national health-care delivery system such as the National Health Service in the United Kingdom. But the need for health-care systems that learn is similar the world over.

## COMBINE CARE AND RESEARCH

Some health-care organizations are working out ways to collect evidence from ongoing care to improve patient outcomes. Such efforts systematically analyse clinical data from peer-reviewed literature and introduce selected findings into clinical care. Then they track and evaluate outcomes.

One of the best examples of such a ‘learning health-care system’ is the Geisinger Health System, which serves 2.6 million patients in Pennsylvania. Once diagnosed with chronic conditions such as diabetes, heart failure or depression, patients’ care and outcomes are monitored. Metrics such as hospitalization rates are used to recommend that certain individuals receive coordinated attention from a team of specialists including mental-health practitioners and

## GENES THAT DOCTORS USE

Genetic variants can determine whether a treatment harms or heals.

Condition	Gene or variant	Action
Cystic fibrosis	<i>CFTR-G551D</i> (plus eight other variants)	Select medication only effective in patients carrying the variants.
HIV	<i>HLA-B*5701</i>	Avoid HIV drug abacavir. Select alternate medication that is less toxic to patient.
Melanoma	<i>BRAF V600E/K</i> and similar variants	Select medication that is only effective in patients carrying the variants.
Lung cancer	<i>EGFR</i>	Select medication that is only effective in patients carrying key mutations.
Family history of breast cancer	<i>BRCA1/BRCA2</i>	Opt for close screening or pre-emptive mastectomy.

nutritionists. Over three years, this reduced hospital admissions by 18% and seemed to reduce spending by 7% (ref. 8).

A learning health-care system depends on electronic medical records (EMRs) — comprehensive, accessible stores of patient data. EMRs can, for instance, allow a health system to identify people with diabetes who have also been prescribed blood-pressure medication, and determine rates of heart or liver failure according to age or gender.

Dozens of health-care systems are adapting to conduct genomics research<sup>9</sup>. Several hospitals, including Children’s Mercy Hospital in Kansas City, Missouri; Baylor College of Medicine in Houston, Texas; and the Medical College of Wisconsin in Milwaukee, are using sequencing to unravel maladies that defy diagnosis, including life-threatening intestinal inflammation in a young child and the probable cause of seizures in another.

The UK Genomics Initiative is sequencing 100,000 people’s genomes and linking the resulting data to the participants’ National Health Service records. In the United States, the care consortium Kaiser Permanente and the University of California, San Francisco, are testing whether a set of genetic variants assessed in 100,000 patients might identify those with higher health risks. Vanderbilt University in Nashville, Tennessee, found<sup>10</sup> that 91% of all patients and 96% of African American patients carried one or more variants implicated in affecting response for five common prescription drugs, including warfarin and simvastatin, a cholesterol-lowering agent.

Such ‘pharmacogenomic’ applications — in which genetic markers are used to fit drugs to patients — are among the most promising areas for collecting evidence during clinical care. In Thailand, about 12% of people have genetic predisposition to Stevens–Johnson syndrome, in which certain medications trigger a blistering, life-threatening rash. The government has sponsored a programme in which any Thai citizen can be genotyped to predict reactions to problematic drugs such as carbamazepine, commonly used to control seizures.

Ramathibodi Hospital in Bangkok provides a health card for patients with risky genetic variants to present to pharmacists, alerting them to provide alternative medications. Whether decreased toxicity merits the use of less-effective drugs is being evaluated.

## INVESTIGATE AND INFORM

Other efforts are making genomic information part of routine care during a visit to the doctor or a hospital stay. The US National Human Genome Research Institute is now funding projects that explicitly explore clinical implementation of genomic information through EMRs. The Electronic Medical Records and Genomics (eMERGE) network, a consortium of nine institutions that use commercial and homegrown EMRs, is developing ways to capture genomic information, store it securely and incorporate it into computer algorithms to guide clinicians.

As more genetic data are incorporated into EMRs, health systems can carry out ‘pragmatic clinical trials’ — those that randomize groups of clinicians to study an intervention (in this case treatments guided by genetic information) compared with usual care. Although standards for how results should be assessed and adopted still need to be worked out, such studies are cheaper than conventional clinical trials and can collect data from people with multiple conditions, who are often excluded from trials.

To translate genetic analysis into improved medical care, hospital executives and other leaders must redesign how they collect and manage data so that it can be used in research. Existing patient databases for clinical care are often inadequate for evaluating health interventions: data are frequently missing or incorrect, and it can be hard to link data about, for example, demographics, appointments, procedures, medications and vital signs.

Pioneers of learning health-care systems have established streamlined consenting processes and data warehouses that link EMRs with genomic assays from patients’ banked biospecimens. However, systems are not consistently incentivized to build these capacities.

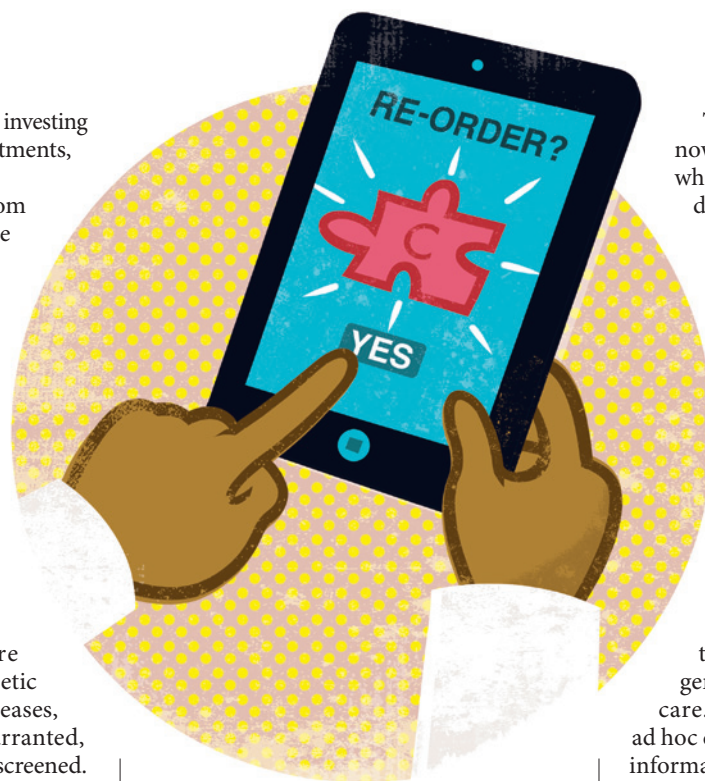


Health leaders want evidence before investing resources, yet without those investments, such evidence is hard to gather.

This makes the experiences from diverse health systems all the more precious. In January this year, a nascent organization of countries and international organizations met to discuss their efforts to evaluate how and whether genetic variants can improve clinical care. Programmes are a blend of data collection (bringing genomic information and medical records together), clinical implementation (getting providers to use information) and outcomes (assessing the impact of using information).

For example, the Singapore National Eye Center is using genetic testing to diagnose some eye diseases, select treatments and, when warranted, encourage family members to be screened. The Luxembourg Centre for Systems Biomedicine and the Centre Hospitalier de Luxembourg are diagnosing and stratifying risk for Parkinson's disease, and the University of Colombo in Sri Lanka is doing the same for inherited blood disorders. Vanderbilt University, the Geisinger Health System and Northwestern University in Evanston, Illinois, are incorporating pharmacogenetic testing to select medications. And Duke University in Durham, North Carolina, allows people to report their family histories of disease online. This flags the individuals most likely to benefit from genetic screens for variants that raise colon-cancer risk or lead to heart defects that can cause sudden death.

Such projects must be sure to provide iterative feedback on how genetic information is used and how it affects outcomes. It is important to understand how and when



physicians and their patients should receive genomic information, and how it is likely to alter their decisions. This will inform when patients should get genetic tests and how complicated information can be delivered to both patients and physicians.

But how will the use of genetic variants in medical decisions move from exploratory to mainstream? The US Centers for Disease Prevention and Control has developed a useful framework to evaluate new genomics tests. It includes whether a test measures genotype accurately, how reliably it provides information about a diagnosis or outcome, how likely it is to improve outcomes and whether the test has ethical, legal or social implications, such as uncovering false paternity or future untreatable conditions.

The medical-genomics community now needs to focus on how to decide what levels of evidence are required to demonstrate that assessing a variant helps patients. For example, a variant that may require a change in diet might need less evidence than one that may require an irreversible surgical procedure. The clinical community should commit to developing ways to evaluate genetic variants as 'actionable', including understanding how results from one patient population might apply to another with differing genetic backgrounds or health practices.

Most health systems around the world are not fit to shepherd genome sciences into routine health care. Each health system still makes ad hoc decisions about how to use genetic information. Holding international meetings and establishing consortia are essential first steps. Hospitals must appreciate that informatics investments will benefit patients and efficiency. Such shifts will be difficult — clinicians may resist, resources will be tight and experiments will fail (see 'Next steps for key players in medical genomics'). Adoption will increase as success stories accumulate and practices change, but those who act early will enjoy some advantages. US systems that made initial investments in institutional biobanks were better placed to compete for federal funding, for example.

Meanwhile research and health-care funders should incentivize those embedding genetic information into clinical care. Providers, regulators, insurers and health ministries will have more confidence in gene-tailored treatments, and patients will hopefully have better care and improved health. ■

**Geoffrey Ginsburg** is director of genomic medicine at Duke University in Durham, North Carolina, USA.  
e-mail: [geoffrey.ginsburg@duke.edu](mailto:geoffrey.ginsburg@duke.edu)

1. Green, R. C. *et al. Genet. Med.* **15**, 565–574 (2013).
2. Lynch, J. A. *et al. Genet. Med.* **15**, 630–638 (2013).
3. Mallal, S. *et al. N. Engl. J. Med.* **358**, 568–579 (2008).
4. Kimmel, S. E. *et al. N. Engl. J. Med.* **369**, 2283–2293 (2013).
5. Pirmohamed, M. *et al. N. Engl. J. Med.* **369**, 2294–2303 (2013).
6. Vogelstein, B. *et al. Science* **339**, 1546–1558 (2013).
7. Green, E. D. *et al. Nature* **470**, 204–213 (2011).
8. Gilfillan, R. J. *et al. Am. J. Manag. Care* **16**, 607–614 (2010).
9. Manolio, T. *et al. Genet. Med.* **15**, 258–267 (2013).
10. Van Driest, S. L. *et al. Clin. Pharmacol. Ther.* **95**, 423–431 (2014).

## TO-DO LIST

### Next steps for key players in medical genomics

- **Health-care providers.** Learn to use electronic medical record (EMR) enabled tools and to apply genomic information to clinical decisions.
- **Patients.** Define preferences about what information they wish to learn and have placed in their EMR.
- **Clinical molecular geneticists.** Interpret how genetic results change biological function, including variants that warrant clinical action.
- **Bioinformaticians and computing specialists.** Develop curated genomic databases and means to query them to guide clinical decisions.
- **Genetic counsellors.** Provide information about the potential burden of results and implications for family members.
- **Policy-makers.** Define the rules to return results to patients and ensure privacy and confidentiality.
- **Researchers.** Develop consensus on how best to conduct research using EMRs linked to genomic information.
- **Translational specialists.** Train people across disciplines to communicate effectively and provide the 'glue' for the system to work.
- **Software vendors.** Design effective EMR and related systems.





© THE ARMITT TRUST

Beatrix Potter's 1895 drawings of the *Boletus granulatus* mushroom were part of a study that she submitted to the UK Royal Botanic Gardens at Kew.

## NATURAL HISTORY

# A scientist's eye

Beatrix Potter's meticulous artistry served mycology and entomology as well as children's fiction, reveals **Linda Lear**.

In January, the British press reported the discovery of a rare parasitic fungus on the Mar Lodge Estate in Aberdeenshire. Liz Holden, an independent field mycologist, spotted the small jelly fungus *Tremella simplex* growing on the pink blobs of another rarity, *Aleurodiscus amorphus*. When she checked, she discovered that *T. simplex* had first been drawn in the late 1890s, by Beatrix Potter (1866–1943).

Before Potter became a famous children's author and illustrator, she was a pioneering naturalist and amateur mycologist, although later discouraged by professionals in Britain's natural-history establishment. It was her habit to draw everything she saw under the lens, so Potter included the *Tremella* in her study, although she could not have recognized it

then as an independent parasitic fungus. Potter was an extraordinary observer whose many contributions to natural science are only now becoming more widely recognized. Along with women such as Margaret Gatty, author of *The History of British Seaweeds* (1863), Potter was part of a generation of female naturalists whose work contributed to the advancement of professional science, whether acknowledged or not.

Potter always prized the tribute paid to her by family friend John Everett Millais, the Pre-Raphaelite society painter: "plenty of people can draw, but you ... have observation". All her life, she exhibited a meticulous

concern for factual evidence. Her recording of observable data, although deliberately never systematic because she followed her artistic inclinations, marked her as a student of natural history from a young age. At nine, she was executing watercolour sketches of caterpillars, complete with physical descriptions and field observations. That she was interested in geology, archaeology, entomology and especially mycology was not unusual for someone raised in wealth and privately educated. What was rare was how Potter used her gifts in diverse areas, from stories for children and animal husbandry to the preservation of land, farms and watersheds in the English Lake District.

Potter's childhood offered unique opportunities for observing and recording

**NATURE.COM**  
For Linda Lear on  
Potter's scientific  
legacy:  
[go.nature.com/zzibm1](http://go.nature.com/zzibm1)

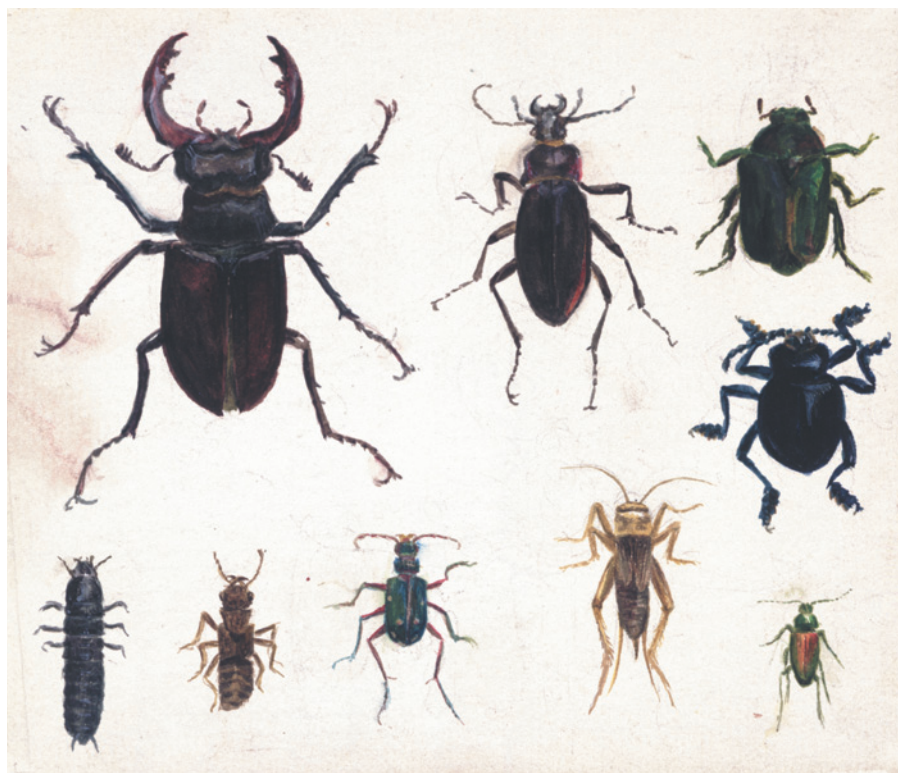
nature. She enjoyed summers exploring and drawing the flora and fauna of Perthshire near the River Tay in central Scotland, tagging along with her artistic parents and absorbing photographic techniques of perspective and detail from her father, a fine amateur photographer. As a young woman, she explored the Tay Valley in her pony and trap, noting in her journal geological formations, diversity of land use and the progress of soil erosion, and despairing over practices such as the dehorning of Ayrshire cattle.

### SCHOOLROOM SCIENCE

The boredom of the Victorian schoolroom enhanced Potter's skills almost by default. After lessons, Beatrix and her younger brother Walter drew a menagerie of animals secretly conveyed into the nursery — rabbits, mice, hedgehogs, bats, snails and lizards — as well as more typical collections of insects and bird eggs. When a schoolroom pet died, the Potter children often boiled the corpse and articulated the bones to improve the anatomical accuracy of their drawings. Potter noticed that lettuce contained a “soporific” that made her pet rabbit Benjamin sleepy; that field mice were inordinately fastidious housekeepers; and that hedgehogs yawned “pathetically” and might bite when propped up in one position to draw. Such discoveries later informed the plots and characters of her children's tales.

Like the artist and critic John Ruskin, Potter understood that the only way to know something was to draw it. First the hand-lens, then the camera, and finally the microscope taught Potter how to ‘see’. By her early 30s, Potter's enthusiasm was focused on how fungal spores reproduced — an issue that few British mycologists agreed on. During a holiday in Scotland in 1892, Potter had formed a botanical alliance with noted naturalist Charles McIntosh, who provided instruction in the microscope drawing of fungi in exchange for Potter's accurate watercolours of rare specimens. By 1895 Potter had gathered young forms of the mushroom *Boletus granulatulus*, now known as *Suillus granulatus*, and drawn the spores and spore-producing structures, or basidia. Potter successfully germinated spores of several species of fungi, and made drawings of the mycelium at different stages.

She approached the Royal Botanic Gardens at Kew with these



A Beatrix Potter study of insects, including a stag beetle (top left) and a bloody-nosed beetle (middle right).

findings, only to be rebuffed by its director, William Thiselton-Dyer. She noted in her journal that “he hadn't the time to look at my drawings”, even though he “indicated the subject was profound”. Her uncle — the chemist Henry Enfield Roscoe — encouraged her to continue her research, and in 1897 she offered to the Linnean Society in London (which did not then admit women, or allow them to attend meetings) a paper: ‘On the Germination of the Spores of *Agaricineae*’, which was accompanied by several of her microscope drawings. Although this paper has been lost, it seems from her drawings and journal that Potter had become intrigued with the possibility of hybridization.

Around the mid-1890s, Caroline Martineau, the principal of London's Morley Memorial College for Working Men and Women, commissioned Potter to produce a dozen lithographs to accompany lectures on entomology. Two survive; one shows a sheetweb spider, *Linyphia triangularis*. They are accurate, despite Potter's recorded frustration with the errors in the Natural History Museum's

entomology index, and the misidentified specimens in the museum's insect cases.

Long after Potter had become the celebrated author of *The Tale of Peter Rabbit*, *The Tale of Two Bad Mice* and more than 20 other classic books for the young, she gave her prized mycological and botanical drawings to the Armitage Museum and Library in Ambleside in the Lake District. Today, they are still consulted by professional and amateur mycologists, and 59 of the drawings are reproduced in W. P. K. Findlay's *Wayside and Woodland Fungi* (Warne, 1967).

When her eyesight diminished, Potter turned to breeding prizewinning native Herdwick sheep, and to promoting the preservation of the unique ecology and farming character of the Lake District. On her death in 1943, Potter, then Mrs William Heelis, bequeathed to the National Trust more than 1,700 hectares of land, now enjoyed by thousands of visitors each year.

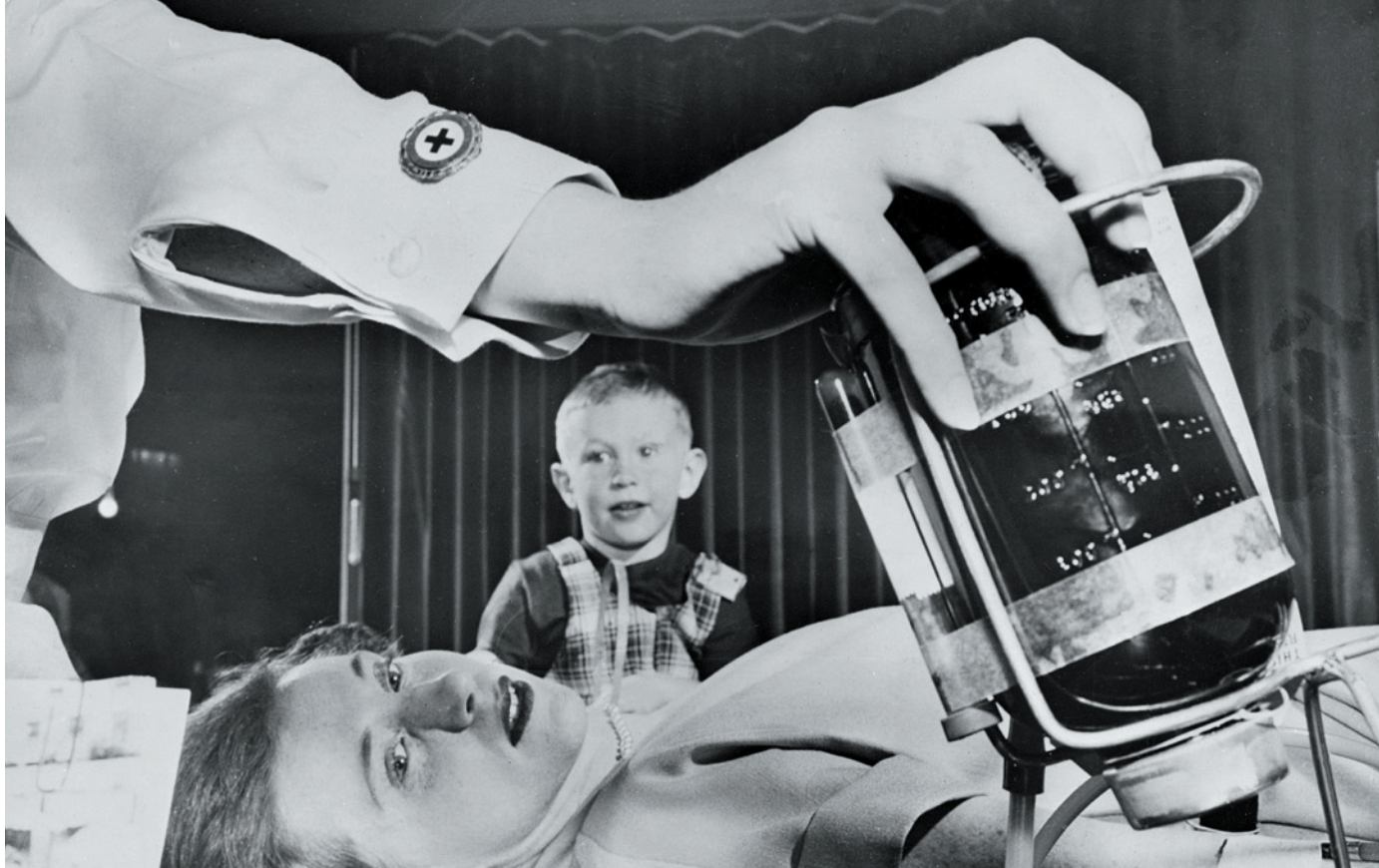
In 1896, Potter summed up her delight in the natural world with a proclamation of supreme Victorian self-confidence: “With opportunity the world is very interesting.” The natural-history legacy of this shy but hugely curious and determined amateur continues to enlighten, and even dazzle. ■

**Linda Lear** is an environmental historian and biographer based in Washington DC. She is the author of *Beatrix Potter: A Life in Nature* (2007) and *Rachel Carson: Witness for Nature* (2009).  
e-mail: [linda@lindalear.com](mailto:linda@lindalear.com)



Beatrix Potter as a teenager, with her spaniel, Spot.





HULTON-DEUTSCH COLLECTION/CORBIS

In the United States of the 1950s, blood donation was at the centre of a debate on the ethics of trading in body tissues.

## MEDICINE

# The commodified body

Scott Carney assesses a study of banked human blood, sperm and milk.

The US medical community's view of the human body shifted in the twentieth century, when it began to be seen as a source of valuable tissues and fluids. In her thought-provoking *Banking on the Body*, Kara Swanson recounts how the use of human blood, semen and breast milk in medicine was made possible only by uniting interlocking and occasionally contradictory systems of body-part or product procurement and storage under the broad concept of 'banking'. The financial metaphor set the stage for a showdown between competing ideas on how the body is exchanged as either a gift or a commodity. Although she focuses on fluids, Swanson argues that the same lessons can be applied to all body parts, including kidneys or faces for transplantation, and human eggs.

Swanson reveals how the body's gradual commodification played out. She recounts how mothers in Boston, Massachusetts, were paid 60 cents a quart for their milk in 1929, but how demand for the fluid declined with the rise of formula milk. Blood was often more complex. At different hospitals, doctors either bought blood or made recipients promise to donate later. At times, the system of tissue exchange reflected contemporary

national dialogue. So in the Second World War, blood was a community resource donated freely to keep soldiers alive. During the cold war, however, creeping fear of communism made "free blood" unpatriotic — a foretaste of the way socialized medicine is seen by many in the United States today. In the 1960s, when for-profit blood banks in Kansas City, Missouri, sued Red Cross-style community blood banks, the arguments hinged on defining whether blood was an interchangeable commodity or a "gift of life". The 'currency' of sperm banking has proved less fungible: sperm's value fluctuates with the donor's genetic profile, with geniuses commanding higher prices.

In her close historical analysis, Swanson, an intellectual-property lawyer, vaults over the profit-or-gift debate by showing that pragmatism reigns. In times of



**Banking on the Body: The Market in Blood, Milk, and Sperm in Modern America**

KARA W. SWANSON  
Harvard University Press: 2014.

shortage, physicians who support free blood exchange have routinely resorted to paying donors to bolster supplies. Blood becomes a commodity when selling it helps to build institutional profitability, as when the American Association of Blood Banks (AABB) lobbied to purchase blood in the 1950s. In the early 1960s, the AABB reversed its position when faced with the possibility of liability lawsuits after a patient received hepatitis-infected blood. Hospitals in New York argued that they were not selling blood — only the medical services to transfer it. Distinctions between products and medical services became hopelessly tangled.

So in today's medical lingo, a person in need of blood in the United States cannot purchase a pint: rather, they rent the expertise of hospitals and surgeons to move the blood from one body to another. The blood is effectively free; the transfusion might end up costing thousands of dollars. The result is that everyone seems to be making money on body parts except for the donors. 'Altruistic donations' are often deemed gifts only while they are being procured. Once alienated from the source, the commodified body parts can be sold as services between



different body banks and, ultimately, to consumers at high prices.

Swanson's solution is simple: scratch the doublespeak and start again. "Body products are property," she writes in her conclusion. "Markets in body products can be harnessed to serve communal goals. The professional donor can be a safe and respected supplier of body products." She makes clear that her aim is to redefine how all tissues — from blood to kidneys — are traded. What she does not quite articulate is that not all body parts are truly equal.

Blood, milk and sperm — and of more recent interest, faeces — are the body's renewables. The removal of a body part, however, entails considerably higher sacrifices. The danger, of course, is that as body products are commodified, so are people. Almost all the voices in her account are doctors, middlemen or regulators; donors and recipients are not represented. We do not hear from prisoners who have been forced to sell their blood, or the many in Pakistan and India who are so desperate for cash that selling a kidney illegally looks like a viable solution.

More importantly, Swanson fails to explore the political economy of the body business. The sale of blood and blood derivatives worldwide generates an annual US\$23.5 billion, and the United States is the single largest blood exporter in the world. The country also exports corneas, coagulants, bones and ligaments. Americans, in turn, purchase kidneys, hearts, eggs and livers, and hire surrogate wombs in foreign markets. As the markets cross borders and profits grow, they render the donors invisible. Medical outcomes are important; so, too, is being sure that the donor is unharmed in the exchange. Looking to the banking structures of the past might not be the best way to create ethical supplies in the future.

For scholars interested in the narrow confines of blood, milk and sperm banks in the United States, Swanson's work is a comprehensive historical sweep. However, without an economic analysis, her examination of banking metaphors offers an incomplete picture of a complex issue. ■

For scholars interested in the narrow confines of blood, milk and sperm banks in the United States, Swanson's work is a comprehensive historical sweep. However, without an economic analysis, her examination of banking metaphors offers an incomplete picture of a complex issue. ■

**Scott Carney** is the author of *The Red Market* (2011) and a senior fellow at the Schuster Institute for Investigative Journalism.

[www.scottcarney.com](http://www.scottcarney.com)

## CLIMATE SCIENCE

# Stratospheric folly

**Tim Kruger** examines an argument against injecting aerosols into the atmosphere to counter climate change.

As evidence of climate change piles up and pessimism grows over climate negotiations, discussion of geo-engineering — deliberate large-scale intervention in Earth's natural systems — is rising up the agenda. In the succinct *Can Science Fix Climate Change?*, climate scientist Mike Hulme focuses on a proposal to cool Earth by injecting aerosols into the stratosphere to reflect solar radiation. He provides a lucid counterpoint to *A Case For Climate Engineering* (MIT Press, 2013) by David Keith, a leading proponent of research into this approach.

Hulme believes that deploying stratospheric aerosols is wrong-headed — undesirable, ungovernable and unreliable. He devotes a chapter to each.

On undesirability, Hulme explains that although the approach would reduce global temperatures, it would also have significant side-effects, such as changing local rainfall patterns. But Hulme compares the downsides with the climate of today, rather than with that of a climate-changed future. This is the equivalent of condemning a drug for having side-effects in healthy people before even considering whether the benefits would outweigh any side-effects in the ill. He thus avoids sully himself with the consideration of what would constitute the lesser of two evils — a climate-changed world without stratospheric aerosols, or one with them.

Hulme is on much stronger ground in looking at whether such approaches are governable. He argues that "If the deployment of the technology cannot conceivably be adequately governed, then the technology itself should not be researched". He places the ball firmly in the court of those proposing research, urging them to explain up front what regulation should consist of. He puts forward three models — a multilateral United Nations process, a consortium-based regime or deployment by a single country — all of which he finds wanting. Least unfavoured would be a multilateral approach, but Hulme questions whether this could deliver, citing the stalled international climate negotiations as evidence of impracticality. A consortium — a coalition of the willing who agree on how to set the global thermostat — he condemns as too flimsy to govern a technology effectively for the requisite decades. And he views deployment by a single country as deeply unsatisfactory.

It is clear that comprehensive governance needs to be in place before deployment, but



**Can Science Fix Climate Change? A Case Against Engineering**  
MIKE HULME  
Polity Press: 2014.

is it necessary before research can be done? Hulme argues that it is, invoking a slippery-slope argument — that once research is started it will lead inexorably to deployment. But although the phenomenon of technological and sociological lock-in is frequently remarked on, so too is the observation that most innovations fail.

Research may show that stratospheric aerosols are far from a satisfactory solution. Hulme argues that their action is unreliable, reeling off a long list of known environmental, social, political and ethical issues, and anticipating others yet unknown. Hulme argues that this list is so overwhelming that it is not worth even starting such an exploration.

Having rejected the case for this technology, Hulme concludes with an alternative plan for tackling climate change. He reiterates the agenda of *The Hartwell Paper: A New Direction For Climate Policy After The Crash Of 2009* (LSE, 2010), calling for "reducing weather risks; improving air quality; innovating in the search for cheap, reliable, clean energy". And he emphasizes the moral character of the challenge we face: "It is not climate change that is the ultimate threat to human well-being. It is the lack of virtue."

The question left hanging is whether Hulme's prescription will be adopted. Would that it were.

One can embrace the Hartwell agenda while recognizing that even if it were to succeed, it would lead to a warming world. And while we await a social tipping point — a global renaissance of humility, compassion and justice — we would be well-served by adopting those virtues in consideration of imperfect remedies. Addressing the knotty problems of climate change is more than an issue of virtue or wickedness. Humanity may yet find itself in the position of having to decide which option is the least worst. ■

**Tim Kruger** manages the Oxford Geoengineering Programme at the University of Oxford, UK.  
e-mail: [tim.kruger@oxfordmartin.ox.ac.uk](mailto:tim.kruger@oxfordmartin.ox.ac.uk)

# Correspondence

## US sanctions put research at risk

Scientific collaboration during the cold war was one of the few links between the United States and the Soviet Union. It is sad and ironic that today's partnerships may now bear the brunt of US sanctions as a result of the situation in Ukraine (see *Nature* **508**, 162; 2014).

Earlier this month, the FBI warned universities in Boston, Massachusetts, about venture-capital partnerships with Russia (see [go.nature.com/oypxph](http://go.nature.com/oypxph)). The US government is also severing scientific exchange by NASA and the Department of Energy with scientists in Russia. These moves are causing widespread consternation in the Russian research community.

Former Soviet and now Russian scientists have traditionally subscribed to democratic and pro-Western agendas (the nuclear physicist Andrei Sakharov is one famous example). Moreover, researchers, engineers and entrepreneurs of Russian origin have contributed significantly to the US enterprise. Indeed, after the collapse of the Soviet Union, the US government issued fast-track green cards to Russian scientists to retain their expertise.

By contrast, the new US policies are likely to be counterproductive. The presumption that Russian scientists are potential spies will force them to return to or stay in Russia, where research is in disarray because of sweeping reforms. Researchers whose projects are derailed are likely to blame the United States rather than Russia — a situation that could further damage relations.

There are signs that Russian scientists with foreign collaborators may come under more state control. It seems that both US and Russian bureaucrats lack understanding of the importance of international scientific collaborations and the way modern science works.

**Konstantin Severinov** *Rutgers*

*University, New Jersey, USA; and Skolkovo Institute of Science and Technology, Russia.*  
[severik@waksman.rutgers.edu](mailto:severik@waksman.rutgers.edu)

## Ecology must seek universal principles

I disagree with your advice to stop searching for universal principles in ecology (*Nature* **507**, 139–140; 2014). It is fortunate that Darwin held no such view. And in this era of big data, big theories are still needed — even in ecology.

The “caveats and exceptions” in existing theory, which you imply scuttle efforts to build better theory, are the very foundation for improvement. This is how science progresses; ecology is no exception.

You recommend that “ecologists should embrace the non-predictive side of their science”, but prediction is part of what distinguishes science from other worthwhile endeavours; it enables us to strengthen and refine our understanding.

Using the wolves in Yellowstone National Park as an example to support your view is ironic. Top-down trophic control in multispecies populations is not a theory: it is an oversimplified conceptual model — a good story that is known to apply sometimes. It remains to be seen whether a unified theory of trophic, competitive and facilitative interactions among species can be achieved, but ecologists should not be advised to give up now.

**John Harte** *University of California, Berkeley, USA.*  
[jharte@berkeley.edu](mailto:jharte@berkeley.edu)

## Mental health: more than neurobiology

The decision by the US National Institute of Mental Health (NIMH) to fund only research into the neurobiological roots of mental disorders (*Nature* **507**, 288; 2014) presumes that these all result from brain abnormalities. But this is not the case for many

people with mental-health issues and we fear that this policy could stall clinical insight into mental illness for years to come.

There is no consistent biological evidence to support the idea that all mental disorders are due to brain dysfunction. Mood and anxiety, for example, are multifactorial and depend on biological, psychological and environmental factors.

The NIMH's assumption of underlying neural mechanisms presupposes that symptoms of specific disorders cluster because they have the same biological cause. However, psychopathological symptoms of disorders such as depression differ dramatically in their causes and genetic context and do not have a common biological background (E. I. Fried *et al. Psychol. Med.* <http://doi.org/r93>; 2013). Symptoms are more likely to cluster because of causal connections — for example, insomnia and its side effects are widespread features of various psychiatric conditions.

**Eiko Fried, Francis Tuerlinckx** *University of Leuven, Belgium.*  
[eiko.fried@gmail.com](mailto:eiko.fried@gmail.com)  
**Denny Borsboom** *University of Amsterdam, the Netherlands.*

## Mental health: drug search on risky path

I fear that the requirement by the US National Institute of Mental Health for investigators to come up with a hypothesis as an initial part of the drug-discovery process (*Nature* **507**, 273 and 288; 2014) will not end well.

The discovery of all early and most recent drugs has hinged on testing the effects of small molecules on the phenotype of cells and organisms (M. J. Keiser *et al. Biochemistry* **49**, 10267–10276; 2010). A hypothesis is not needed (D. B. Kell and S. G. Oliver *BioEssays* **26**, 99–105; 2004) and finding a drug's likely targets can be postponed until it is known whether the drug works.

It is now more fashionable to start the drug-discovery process with a hypothesis on how a drug might act on (typically) a single target, then to test how effective it is. But if the hypothesis turns out to be wrong, which is not unusual (D. B. Kell *FEBS J.* **280**, 5957–5980; 2013), money and effort will have been wasted.

**Douglas Kell** *University of Manchester, UK.*  
[dbk@manchester.ac.uk](mailto:dbk@manchester.ac.uk)

## Restore public trust in care.data project

Public confidence has already been undermined in the National Health Service's care.data programme in England, which will integrate medical records into a database for health-care and research purposes (*Nature* **507**, 7; 2014). We suggest that the UK government's decision to limit release of the data should be the start of a process to develop public trust in large-scale health research, not an end to it.

Care.data first needs to provide more information to the public about how medical records will be used. Openness and clarity are essential. For example, the UK Health and Social Care Information Centre (which manages the data) can currently share pseudonymized data with ‘such persons and in such form and manner... as it considers appropriate’: these choices should be fully explained to the public.

A proposed amendment to the Health and Social Care Act 2012 might narrow this discretion so that data can be shared only for ‘the promotion of health’. But even this is open to wide interpretation.

Formal regulation is only one way to reassure the public. Another is to use online interfaces that allow research participants to access information and to set consent preferences (J. Kaye *et al. Eur. J. Hum. Genet.*, in the press).

**Colin Mitchell, Linda Briceno Moraes, Jane Kaye** *University of Oxford, UK.*  
[jane.kaye@law.ox.ac.uk](mailto:jane.kaye@law.ox.ac.uk)



## Catalysis at the boundaries

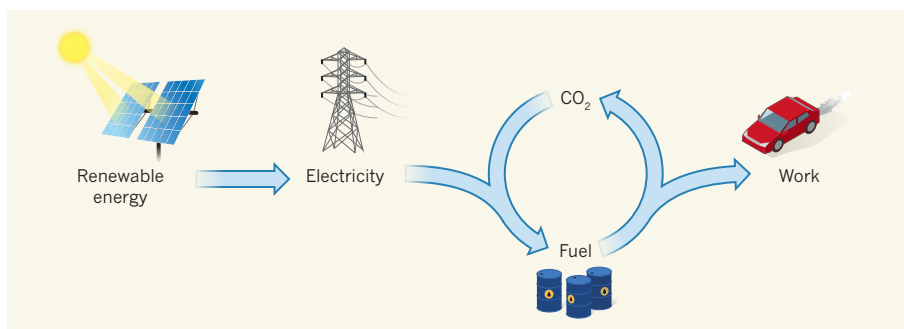
Copper-based materials have been found that efficiently convert carbon monoxide and water to ethanol using electricity. The discovery is a major advance towards storing renewable energy in the form of a liquid fuel. [SEE LETTER P.504](#)

AARON M. APPEL

A remarkable improvement in a catalyst for the electrochemical production of carbon-based fuels from carbon monoxide and water is reported by Li *et al.*<sup>1</sup> on page 504 of this issue. Although electrodes made from copper have been known for many years to catalyse the transformation of carbon dioxide to chemicals and fuels<sup>2</sup>, through intermediates such as CO, the authors have demonstrated that the method used to prepare copper electrodes has a substantial effect on the activity and energy efficiency of these catalysts. This result may lead to improvements in how renewable energy sources are exploited.

Renewable sources such as sunlight and wind provide an opportunity to power society without the negative consequences that accompany the use of fossil fuels, but their intermittent availability is a major limitation. Efficient means of storing energy from these sources must therefore be found in order to facilitate their widespread use<sup>3</sup>. An attractive approach is the use of electricity to drive the production of fuels from water and CO<sub>2</sub> (refs 3–6; Fig. 1). Unlike conventional, centralized fuel production, electrochemical systems can operate at mild pressures and temperatures in small-scale reactors — making them ideal for producing fuels at sites of renewable energy sources, which are inherently dispersed. However, the development of efficient catalysts will be essential for the intended chemical transformations.

The electrochemical conversion of CO<sub>2</sub> into fuels is a multi-step process that has many potential products and intermediates. The choice of catalyst has a profound effect on the selectivity and energy efficiency of this process. The first product is often CO: this gas has a low energy density, so further transformation is required to make an effective fuel. Although copper catalysts are known to work for the initial reduction of CO<sub>2</sub> (refs 2, 7), Li and colleagues focused on the subsequent transformation of CO. They report that, when copper is oxidized and then reduced back to copper in a controlled way, the resulting catalytic metal surface not only produces ethanol more selectively than previous catalysts, but is also more energy efficient. This discovery is a great step towards the cost-effective



**Figure 1 | Storing energy from renewable sources.** Renewable energy sources, such as solar energy, are only intermittently available, so a means of storing the energy is needed. If the energy is converted into electricity, this could be used to drive electrochemical reactions that convert carbon dioxide into liquid fuel, often with carbon monoxide (not shown) as an intermediate; when the fuel is later used to do work, CO<sub>2</sub> is released, and can be recycled into fuel. Li *et al.*<sup>1</sup> report improved copper catalysts for the electrochemical conversion of CO to ethanol, a potential liquid fuel.

production of renewable liquid fuels with high energy densities.

One of the long-standing challenges in heterogeneous catalysis, in which reactions typically occur at the surface of a material, is to understand the causes of changes in catalytic performance. The surface of a material usually bears a variety of structural features, which are potential catalytic sites for the key reactions. Identifying which of these is responsible for catalysis requires the characterization of materials by numerous techniques, followed by attempts to correlate changes in surface features with changes in catalytic performance.

Li *et al.* analysed the copper catalysts produced by their oxidation–reduction process, and found that the size of the particles formed does not seem to explain their catalytic activity — similarly sized copper particles prepared using a different method did not have similar catalytic activity. The researchers propose that the different activities of the catalysts produced using the two methods derive from grain boundaries, the junctions between crystals within the particles. Although particles prepared using different methods may be similar in size, their grain boundaries vary substantially. These multidimensional ‘defects’ may be the key to the significant enhancements seen in selectivity and energy efficiency.

The insight gained from Li and co-workers’ results lays the groundwork for further advances. As with all electrocatalysts, the

energy efficiency of ethanol formation in the authors’ system could be improved by decreasing the electric potential required for reaction and by increasing the product selectivity. But a more remarkable advance may come from studying the carbon–carbon bond-forming process that connects CO molecules to form two-carbon molecules (such as ethanol). The selective formation of carbon–carbon bonds is a key step in a variety of catalytic transformations, not only for industrial applications, but also in biological systems. Enzymes contain precisely positioned chemical groups that control bond-forming and bond-breaking reactions<sup>8</sup>, and the variations in activity and product selectivity observed by Li *et al.* could be due to the formation of similarly multifunctional catalyst structures. An understanding of this process could lead to the production of fuels that contain more carbon atoms (such as butanol), which have an even greater energy density than ethanol.

Liquid fuels produced from CO<sub>2</sub> and its derivative CO face considerable challenges in the fuel market. One of the biggest problems is that fossil fuels have an inherent advantage, because energy was stored in them — by prehistoric photosynthesis — at no cost. By contrast, the energy used to produce fuels from renewable sources must be paid for. At present, this expense is substantial, but the cost relative to that of fossil fuels will probably decrease with time. Because of the challenges of

predicting fossil-fuel prices, renewable-energy production costs, and incentives (such as carbon taxes) to use renewables in place of fossil fuels, the timetable for widespread adoption of renewable fuels is not clear. But the production of fuels from non-fossil sources will certainly require effective catalysts. Li and colleagues' work is an excellent step in that direction. ■

Aaron M. Appel is in the Catalysis Science

Group, Pacific Northwest National Laboratory, Richland, Washington 99352, USA.  
e-mail: aaron.appel@pnnl.gov

1. Li, C. W., Ciston, J. & Kanan, M. W. *Nature* **508**, 504–507 (2014).
2. Hori, Y. in *Modern Aspects of Electrochemistry* Vol. 42 (eds Vayenas, C. G. et al.) 89–189 (Springer, 2008).
3. Cook, T. R. et al. *Chem. Rev.* **110**, 6474–6502 (2010).
4. Olah, G. A., Prakash, G. K. S. & Goepfert, A. J. *Am.*

- Chem. Soc.* **133**, 12881–12898 (2011).
5. Kondratenko, E. V., Mul, G., Baltrusaitis, J., Larrazábal, G. O. & Pérez-Ramírez, J. *Energy Environ. Sci.* **6**, 3112–3135 (2013).
  6. Thoi, V. S., Sun, Y., Long, J. R. & Chang, C. J. *Chem. Soc. Rev.* **42**, 2388–2400 (2013).
  7. Li, C. W. & Kanan, M. W. *J. Am. Chem. Soc.* **134**, 7231–7234 (2012).
  8. Appel, A. M. et al. *Chem. Rev.* **113**, 6621–6658 (2013).

This article was published online on 9 April 2014.

## SENSORY SYSTEMS

# Do you hear what I see?

**Researchers have found evidence that the representation of auditory and tactile information in the brains of blind people shows strong similarities to the way in which visual information is represented in sighted people.**

IONE FINE

Almost one-quarter of the brain is normally devoted to processing visual information: reading text, recognizing faces, following the Sunday match, and much more. The brain's visual cortex contains specialized regions devoted to processing motion, text, faces, scenes, objects and even the position and movement of bodies. In congenitally blind individuals, much of the 'visual' cortex responds strongly to auditory and tactile input rather than to visual stimuli, a phenomenon known as cross-modal plasticity. Writing in *Current Biology*, Striem-Amit and Amedi<sup>1</sup> shed light on how these cross-modal responses are organized.

The researchers trained congenitally blind individuals for 70 hours in the use of a technology<sup>2</sup> that converts visual information into 'soundscapes' (called vOICe). After an initial click, an image frame is scanned from left to right and then represented by a sequence of 'chords'. Brightness is represented by loudness, height in the scene is represented by pitch and horizontal position is represented by the time since the click. For example, a diagonal line stretching upward from left to right becomes the sound of an ascending sweep.

The authors used functional magnetic resonance imaging to measure the brain responses of participants while they listened to soundscapes of body-shape silhouettes, objects, faces and abstract patterns. These brain responses were compared with those of sighted individuals viewing the corresponding visual images. Both blind and sighted subjects showed larger responses to body-shape silhouettes than to other images within

a region of the brain known as the extrastriate body area (EBA), which, in individuals with normal sight, is involved in visual perception of the human body.

One of the most interesting conclusions that Striem-Amit and Amedi draw from their results is that the specialization for body-shape

information in the EBA is innate, requiring little experience to develop, because selectivity for body-shape silhouettes in this brain area occurs even in blind individuals with little or no experience of body shape before training with vOICe. Thus, the authors interpret their findings as evidence that the specialization found in the visual cortex may exist independently of experience.

This claim is based on the assumption that congenitally blind individuals have a deeply impoverished experience of others' body postures, actions and movements. Indeed, the authors claim that this information is available to blind individuals solely through touch, and that they have no auditory experience of body shape before vOICe training. However, in the course of our research, my laboratory

has noticed that many visually impaired individuals report knowing a surprising amount about body positions, actions and intents from sound alone. Visually impaired subject Mike May reports: "I know when people are getting bored by my presentation because I can hear them shuffling in their chairs." A fully blind colleague, Nick Giudice, admits: "When I'm interested in a woman, I can tell if she's slender by how much the floor creaks as she walks by." Amy Burk, the Canadian goalball player shown on the right in Figure 1, reports: "I can see it all ... I can visualize it all ... if they are moving with the ball I'm moving with them ... whatever side they are going on I'm following the person." It is reasonable to suggest that the EBA has a role in interpreting these auditory experiences in blind individuals.

Regardless of whether auditory experience before training with vOICe involves the EBA, this study adds to a body of work suggesting that 'functional constancy' is one of the organizational principles underlying cross-modal plasticity in blind people<sup>3</sup>. According to this principle, specialized brain regions continue to serve the same function in blind individuals, but there is a shift in the regions' primary sensory input from sight to hearing or touch. A variety of studies have provided evidence for functional constancy in blind individuals. These showed that the visual-motion



**Figure 1 | Using sound instead of sight.** In goalball, a sport designed for blind athletes, players use sound to determine the position of both the ball and the other players. Striem-Amit and Amedi<sup>1</sup> report that auditory stimuli are processed in blind people by a region of the brain that is associated with vision in sighted individuals.



area of the brain responds to auditory and tactile motion<sup>4–6</sup>, that sounds made by objects are represented in brain regions associated with visual-object recognition<sup>7,8</sup>, and that reading Braille elicits brain responses in the visual word-form area<sup>9</sup>.

There is an appealing elegance to the idea that the representation of auditory and tactile information in the ‘visual’ cortex of blind individuals is analogous to that of sighted people. But the extent of the apparent similarity may be partly a consequence of hypotheses being tested from a sighted perspective. Experiments published so far have relied on measuring brain responses to auditory stimuli that have been experimentally selected and categorized on the basis of what makes intuitive sense to us as sighted scientists. This may result in an overemphasis of apparent similarities.

More-detailed examination of the response profiles of seemingly analogous areas of the brain using a wider variety of stimuli may yet reveal important differences between blind and sighted individuals, both in the way that information is represented and in the function of these areas with respect to perception and behaviour. Indeed, in their discussion, Striem-Amit and Amedi note that the representation of bodies is not identical in blind and sighted individuals. A second region of the brain that seems to represent body shape, the fusiform body area (which may contain a more holistic representation of body position), showed reduced selectivity to body-shape information in the blind individuals they studied. Moreover, the intriguing possibility that blind people have specialized regions that contain representations of the world with no sighted equivalent remains almost entirely unexplored. ■

**Ione Fine** is in the Department of Psychology, University of Washington, Seattle, Washington 98195, USA.

e-mail: [ionefine@uw.edu](mailto:ionefine@uw.edu)

1. Striem-Amit, E. & Amedi, A. *Curr. Biol.* **24**, 687–692 (2014).
2. Meijer, P. B. L. *IEEE Trans. Biomed. Eng.* **39**, 112–121 (1992).
3. Pascual-Leone, A. & Hamilton, R. in *Progress in Brain Research* Vol. 134 (eds Casanova, C. & Ptito, M.) Ch. 27, 427–445 (Elsevier, 2001).
4. Poirier, C. *et al. Neurolmage* **31**, 279–285 (2006).
5. Saenz, M., Lewis, L. B., Huth, A. G., Fine, I. & Koch, C. *J. Neurosci.* **28**, 5141–5148 (2008).
6. Bedny, M., Konkle, T., Pelphrey, K., Saxe, R. & Pascual-Leone, A. *Curr. Biol.* **20**, 1900–1906 (2010).
7. Mahon, B. Z., Anzellotti, S., Schwarzbach, J., Zampini, M. & Caramazza, A. *Neuron* **63**, 397–405 (2009).
8. Amedi, A. *et al. Nature Neurosci.* **10**, 687–689 (2007).
9. Striem-Amit, E., Cohen, L., Dehaene, S. & Amedi, A. *Neuron* **76**, 640–652 (2012).

## INFECTION BIOLOGY

# Nibbled to death

**Trogocytosis, a process in which one cell ‘takes a bite’ out of another, had previously been seen only in immune cells. But the phenomenon has now been found in *Entamoeba histolytica*, as a way for this parasite to kill host cells. SEE LETTER P.526**

NANCY GUILLÉN

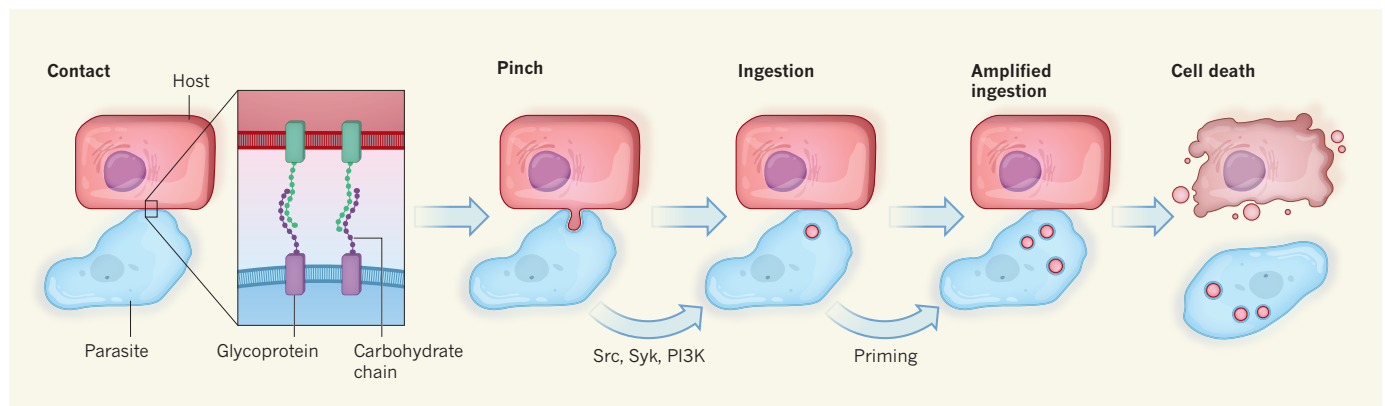
Early in an immune response, antigen molecules are captured by antigen-presenting cells and displayed on the cell’s surface, where they are recognized by receptors on the surface of lymphocytes (B cells, T cells and natural killer cells)<sup>1,2</sup>. The lymphocyte then becomes joined to the antigen-presenting cell and extracts these surface molecules along with patches of the cell

membrane. This process, termed trogocytosis (from the Greek word *trogo*, to nibble)<sup>2,3</sup>, activates the lymphocyte to initiate a specific immune response to that antigen. Until now, trogocytosis had been observed only between immune cells. But in this issue, Ralston *et al.*<sup>4</sup> (page 526) describe a form of trogocytosis carried out by the parasite *Entamoeba histolytica*, and suggest that this process mediates the destruction of intestinal cells that is seen in amoebiasis — the gastrointestinal infection

caused by these unicellular organisms.

Trogocytosis in immune cells requires the transduction of signals from the acceptor-cell surface by means of kinase enzymes such as Src, Syk and PI3K, and by modulation of the cell’s cytoskeleton (which is rich in the protein actin) and of intracellular calcium-ion ( $\text{Ca}^{2+}$ ) levels<sup>2,5</sup>. It is a rapid process, occurring within minutes of co-culturing the participant cells *in vitro*<sup>2</sup>. Notably, despite the exchange of material, neither the antigen-presenting cell nor the lymphocyte dies following trogocytosis.

In the first description of the process<sup>3</sup>, it was proposed that trogocytosis may have evolved as a way for cells to acquire nourishment from other cells, and later as a means of intercellular communication. The hypothesis of an ancient origin for trogocytosis is now supported by Ralston and colleagues’ observation of a similar mechanism in *E. histolytica*, an ancient organism. However, unexpectedly, this form of trogocytosis enhances the parasite’s virulence,



**Figure 1 | Amoebic trogocytosis.** Ralston *et al.*<sup>4</sup> have described trogocytosis by *Entamoeba histolytica*, in which the parasites tear off and ingest patches of host cells, resulting in their death. The interaction between the cells is mediated by abundant surface molecules, including glycoproteins and their attached carbohydrate chains. The cell-to-cell contact is then stabilized through adhesion molecules (not shown), and cytoskeletal activity in the

parasite is probably involved in generating the force required to pinch off the host-cell membrane. This transfer of cellular material causes changes in intracellular calcium-ion levels and triggers signalling pathways in both cells. The activity of Src, Syk and PI3K enzymes contributes to this process and leads to ‘priming’ of the parasites, which amplifies subsequent ingestion. After successive nibbling events, the host cell dies and the parasite moves on.

and results in killing of the target cells.

During amoebiasis, *E. histolytica* resides in the colon of an infected individual, where it depletes it of mucus, interacts with the exposed enterocyte cells lining the colon, dismantles the junctions between them and causes their death by lysis. The previous model<sup>6</sup> for the parasite's action was that it attaches itself to host cells and kills them, by an as-yet unclear mechanism involving the insertion of 'amoebapore' peptides into the cell membrane and subsequent lysis. It was also thought that the parasites engulf and ingest dying enterocytes by phagocytosis. But Ralston *et al.* instead show that the parasite ingests pieces of the host cell, and that this nibbling occurs in a repeated manner that ends up killing the cell. Once killing is achieved, the amoebae move on (Fig. 1).

This trogocytosis-like process is fundamentally different from amoebic phagocytosis, in which the parasite entirely ingests cells such as red blood cells (including dead cells). Although it is not known what determines the parasite's choice between phagocytosis and trogocytosis, some of the authors' findings point to mechanisms similar to those suggested in immune cells<sup>7</sup> that can perform trogocytosis in parallel with phagocytosis. For instance, both processes require an active actin-rich cytoskeleton and associated signalling pathways.

Interestingly, the authors' data show that parasites that have previously performed trogocytosis are 'primed' for higher ingestion activity and elicit more host-cell killing than parasites that have not, demonstrating that trogocytosis changes the parasites' behaviour. The authors have also produced images of trogocytosis in living tissue for the first time, and show that the process is required for the parasites to invade the tissue, leading to the pathogenic consequences for the host. They further show that the parasites can trogocytose all of the cell types tested, including enterocytes, lymphocytes, intestinal-tissue cells and red blood cells; because these cells have different surface constituents, it seems likely that a ubiquitous interaction occurs at the cell surfaces during trogocytosis.

It can thus be argued that the first step in amoebic trogocytosis — cell-to-cell attachment — is mediated by general components of the cell surface (for example, the glycocalyx, which is rich in glycoproteins and glycolipids), rather than by specific cell receptors such as the T-cell receptor (TCR), as is observed in immune-cell trogocytosis. The most relevant glycosylated (carbohydrate-containing) components on the parasite cell surface, in terms of abundance, are lipopeptidophosphoglycans and Gal/GalNAc lectin<sup>6</sup>. Glycosylated residues on these amoebic components might become attached to glycosylated components on the donor cells. The data obtained by Ralston *et al.* support this hypothesis, because Gal/GalNAc lectin is essential for the process.

Despite their respective cell specificity and

ubiquity, there are interesting correlations between the TCR and Gal/GalNAc lectin in terms of signal-transduction characteristics: antigens bind to the TCR with relatively low affinity, similar to the predicted weak interactions between glycosylated residues and the Gal/GalNAc lectin; the activation of both intracellular signalling pathways involves the dynamic linking of the cell-surface molecules into microclusters<sup>8,9</sup> and requires a short cytoplasmic domain; and both pathways are associated with Src-kinase activity in the acceptor cell. However, these are only correlated features, because there is no evidence for structural homology between the TCR and Gal/GalNAc lectin.

It is also interesting to speculate on the significance of the intensification of amoebic trogocytosis in primed parasites. One possible explanation is that surface components of the host cell activate specific signalling pathways that enhance the parasite's affinity for extracellular carbohydrates. This idea is based on recent findings indicating that carbohydrate metabolism is involved in amoebic pathogenesis<sup>10</sup>. Are these signals necessary for killing donor cells by amoebic trogocytosis, but not for lymphocytic trogocytosis? Comprehensive comparisons of gene expression between primed and naive parasites (and in lymphocytes) might help to explain the roles

of these activated signals. Further studies may also reveal whether amoebic trogocytosis and phagocytosis occur simultaneously, and what specific signals result from each process. But although such details remain to be determined, this new concept of amoebic trogocytosis is important for understanding not only amoebiasis, but also host–pathogen interactions in other systems and in immune-cell function and interactions. ■

Nancy Guillén is in the Department of Cell Biology and Infection, Institut Pasteur, Paris 75015, France.  
e-mail: nguillen@pasteur.fr

1. Hudrisier, D., Riond, J., Mazarguil, H., Gairin, J. E. & Joly, E. *J. Immunol.* **166**, 3645–3649 (2001).
2. Davis, D. M. *Nature Rev. Immunol.* **7**, 238–243 (2007).
3. Joly, E. & Hudrisier, D. *Nature Immunol.* **4**, 815 (2003).
4. Ralston, K. S. *et al. Nature* **508**, 526–530 (2014).
5. Aucher, A., Magdeleine, E., Joly, E. & Hudrisier, D. *Blood* **111**, 5621–5628 (2008).
6. Faust, D. M. & Guillén, N. *Microbes Infect.* **14**, 1428–1441 (2012).
7. Daubeuf, S. *et al. J. Immunol.* **84**, 1897–1908 (2010).
8. Chen, L. & Flies, D. B. *Nature Rev. Immunol.* **13**, 227–242 (2013).
9. Blazquez, S. *et al. Cell. Microbiol.* **10**, 1676–1686 (2008).
10. Thibaux, R. *et al. PLoS Pathog.* **9**, e1003824 (2013).

This article was published online on 9 April 2014.

## GENETICS

# The vital Y chromosome

Comparisons of Y-chromosome sequences in various mammals reveal abundant gene loss early in the chromosome's evolution but remarkable gene stability across the Y chromosomes of extant species. [SEE ARTICLES P.488 & P.494](#)

ANDREW G. CLARK

**T**he evolution of the mammalian X and Y sex-determining chromosomes from ancestral chromosomes is thought to have occurred through a rapid loss of genes from the Y chromosome. This idea of rapid degeneration<sup>1</sup> has been bolstered by observations made during the emergence of new Y chromosomes or Y chromosome segments in fruit flies<sup>2</sup>. In this issue, Bellott *et al.*<sup>3</sup> (page 494) and Cortez *et al.*<sup>4</sup> (page 488) present extensive accounts of gene evolution on the Y chromosome. They show that, although there was a period of rapid degeneration and gene loss during its early evolution, the genes that are conserved across the Y chromosomes of extant mammals (and the sex-determining W chromosomes of birds) have since been

remarkably stable. The researchers' data also provide a detailed picture of the evolutionary forces acting on the sex chromosomes, and offer a plausible explanation for the functional coherence of Y-linked genes across these species.

The Y chromosome is notoriously challenging to study, in terms of both genetics and molecular biology. Despite the fact that male genomes were included in early whole-genome sequencing projects, the Y chromosome was largely ignored owing to the challenge of obtaining useful data from the chromosome, which is rich in repetitive and palindromic sequences. Bellott *et al.* adopted a previously described approach<sup>5</sup> — cloning regions of the DNA of interest into bacterial artificial chromosomes — to obtain and assemble DNA sequences from the Y chromosomes





## 50 Years Ago

'Extracorporeal perfusion of the isolated head of a dog' — Critical evaluation of cerebral metabolism and intracranial fluid distribution necessitates complete isolation of the brain's blood supply; however, brain viability must be demonstrated and maintained for such studies to be meaningful ... In order to minimize handling of the brain substance, a factor which may disturb fluid distribution and cerebral metabolism, we have chosen to leave the brain within the skull during perfusion ... Most cortical activity ceased when blood glucose was depleted ... Even after the electrocortical activity ceased, corneal and lid reflexes remained intact and the oxygen and glucose consumption continued ... From our experience, we believe that electrocortical activity is a sensitive index of brain viability, in that it is lost long before inactivation of corneal and lid reflexes or cessation of metabolism. In this preparation we have demonstrated that the dog brain maintains this activity for several hours after complete decapitation.

**From *Nature* 25 April 1964**

## 100 Years Ago

The second reading of a Bill to prohibit experiments on dogs was carried in the House of Commons on Friday last, April 17, by a majority of forty-two ... It was stated on behalf of the Government that an amendment will be moved in Committee to abolish the proposed prohibition and to allow experiments only in cases where no other animal but a dog is available for the purpose ... Before the second reading was taken, a memorial signed by more than three hundred eminent physicians, surgeons, and other scientific investigators, protesting against the measure, was addressed to the Home Secretary.

**From *Nature* 23 April 1914**

of four placental mammals (rat, mouse, bull and marmoset) and the marsupial opossum. They compared these with existing sequences for another three placentals (rhesus macaque, chimpanzee and human). Of the 184 genes that the authors infer to have been on the ancestral sex chromosomes some 300 million years ago, they find that only 3% survive on the Y chromosome of one or more of these mammals (Fig. 1).

Consistent with previous reports, this means that massive degeneration and gene loss did occur early in the history of the mammalian Y chromosome. However, once the genes had run this gauntlet, those that remained enjoyed remarkable stability on the Y chromosome. The authors also find that the 36 genes that are present on both the X and the Y chromosomes of all eight species they examined have maintained a stable presence for the past 25 million years. Ten genes were found to be shared across the Y chromosomes of the tammar wallaby, the Tasmanian devil and the opossum, indicating a stable Y-chromosome presence for the 78 million years of the marsupial lineage. These findings have important implications for our understanding of how natural selection acts to retain active functioning of specific subsets of genes on the Y chromosome.

Cortez *et al.* took a faster survey approach, in which they sought RNA molecules that are expressed in males but not females and then verified that the genes encoding these RNAs are found only in male genomic DNA. This allowed them to identify 134 genes transcribed from the Y chromosome across 10 mammals and to follow their evolutionary fates. By including the chicken (in which males have

two Z chromosomes and females have one Z and one W chromosome) and the platypus (a monotreme that has a bizarre array of five X and five Y chromosomes), the authors were able to paint a broader picture of sex-chromosome evolution. Most noteworthy is their observation that the sex chromosomes of placental mammals, birds and monotremes had essentially independent origins, which means that patterns of gene loss and of specific retention of classes of genes on their Y (or W) chromosomes can be compared.

These data add depth and confidence to the model of evolutionary 'strata' on the sex chromosomes<sup>6</sup> that mark the time points at which X and Y sequences ceased recombining and subsequently diverged. Intriguingly, despite their independent origins, the authors find that the oldest strata in placental mammals, monotremes and birds are remarkably similar in age, estimated to have occurred 181 million, 175 million and 137 million years ago, respectively.

Another key aspect of genes on sex chromosomes is dosage sensitivity. Dosage-insensitive genes are those that function perfectly well when present as a single copy, and these are especially likely to become X- or Y-specific. By contrast, two copies of dosage-sensitive genes are required for normal health, and such genes are likely to be retained on both the X and the Y chromosome<sup>7</sup>. Genes involved in regulating gene transcription — such as those that encode transcription factors — commonly function inadequately in only a single dose, providing a hypothesis for why the Y chromosome has retained genes involved in transcription regulation.



**Figure 1 | Small but stable.** The human Y chromosome (right) is much smaller than the X chromosome (left), as a result of extensive degeneration early in Y-chromosome evolution. However, comparisons with other mammalian Y chromosomes by Bellott *et al.*<sup>3</sup> and Cortez *et al.*<sup>4</sup> show that there has been remarkable gene stability across Y chromosomes following this initial gene loss.

POWER AND SYRED/SPL

Because the Y chromosome is enriched with transcription-regulating genes, this means that it is far from being solely a male-determining switch that is flipped early in development. Instead, the Y chromosome has an impact on gene regulation across the genome in males, potentially influencing biological functions throughout life and in every tissue. It is fair to say that we are only beginning to understand the full extent of the differences in the molecular biology of males and females, and unanswered questions abound. For example, to what extent are male–female differences driven by specific interactions with Y-chromosomal factors?

In humans, the level of variation between

individuals is considerably lower on the Y chromosome than on other chromosomes. However, Y-linked sequence changes can cause changes in gene expression across the genome, which could result in amplified differences among males. Despite the relative stability of the gene content on mature Y chromosomes, it is well known that DNA sequences evolve faster on the Y chromosome than on the X. Although this is generally perceived to be the result of the arrest of genetic recombination on the Y chromosome leading to reduced effectiveness of natural selection<sup>8</sup>, it seems that the Y chromosome also has the potential to mediate remarkably rapid adaptive evolutionary change. ■

**Andrew G. Clark** is in the Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA.

e-mail: [ac347@cornell.edu](mailto:ac347@cornell.edu)

1. Charlesworth, B. & Charlesworth, D. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
2. Bachtrög, D. *Nature Rev. Genet.* **14**, 113–124 (2013).
3. Bellott, D. W. *et al. Nature* **508**, 494–499 (2014).
4. Cortez, D. *et al. Nature* **508**, 488–493 (2014).
5. Skaletsky, H. *et al. Nature* **423**, 825–837 (2003).
6. Lahn, B. T. & Page, D. C. *Science* **286**, 964–967 (1999).
7. Wilson Sayres, M. A. & Makova, K. D. *Mol. Biol. Evol.* **30**, 781–787 (2013).
8. Wilson Sayres, M. A., Lohmueller, K. E. & Nielsen, R. *PLoS Genet.* **10**, e1004064 (2014).

## CLIMATE SCIENCE

# Sea levels from ancient seashells

**The isotopic composition of oxygen in sea water correlates with changes in global mean sea level. Microfossils carrying oxygen–isotope signals have been used to extend sea-level records as far back as 5 million years ago. [SEE ARTICLE P.477](#)**

RALPH SCHNEIDER

On page 477 of this issue, Rohling *et al.*<sup>1</sup> present a convincing approach for calculating sea-level fluctuations over the past 5 million years (Myr). Their method depends on variations in the oxygen–isotope composition of shells produced by unicellular organisms called planktonic foraminifera (Fig. 1). In this way, the authors provide much-needed information that should help to predict future rates of sea-level rise in the event of complete or partial melting of the ice caps over Greenland and Antarctica in response to global warming.

The most recent instances of ice-sheet growth and melting that generated sea-level variation of several metres took longer than hundreds of years, making it impossible to determine their effects directly from historical records. To infer how growing and melting continental ice sheets affect sea level, at least the past 500,000 years must be considered. During this time, there were five periods of sea-level rise of up to 100 m or more, corresponding to the terminations of recent glacial periods (the intervals of time within the current, ongoing ice age that, in general, correspond to colder temperatures and glacier advances).

One could argue, however, that these terminations do not adequately describe what would happen during complete melting of the Greenland ice cap and parts of the West Antarctic

Ice Sheet. Hence, it is much more appropriate to consider past conditions when Northern Hemisphere glaciation was still young — that is, when small, juvenile continental ice caps controlled sea-level fluctuations on timescales of tens to hundreds of thousands of years (the timescales associated with variations in Earth's orbit, which dominate the timing of glacial–interglacial cycles and sea-level changes). But how can this be done? The shells of foraminifera offer a potential solution. Made of calcium carbonate, they contain a record of the ambient isotopic composition of seawater oxygen during the organisms' lifetime.

Since the pioneering work<sup>2</sup> of the geologist Cesare Emiliani in the 1950s, it has commonly been accepted that periodic variations in the ratios of oxygen-18 to oxygen-16 ( $^{18}\text{O}/^{16}\text{O}$ ) in foraminifera preserved in deep-sea sediments follow a global pattern characteristic of orbitally forced climate change during the Late Pleistocene epoch (about 700,000 to 11,700 years ago). Cool temperatures and great ice volumes both resulted in high  $^{18}\text{O}/^{16}\text{O}$  ratios, whereas high temperatures and low ice volumes had the opposite effect. So, if the temperature effect can be disentangled from this isotopic record, then the remaining signal represents relative changes in continental ice volume. And if this signal can then be scaled to the amplitude of sea-level rise between glacial and interglacial periods (as has been done for the most recent postglacial period by correlating sea-level rises to the oxygen–isotope com-

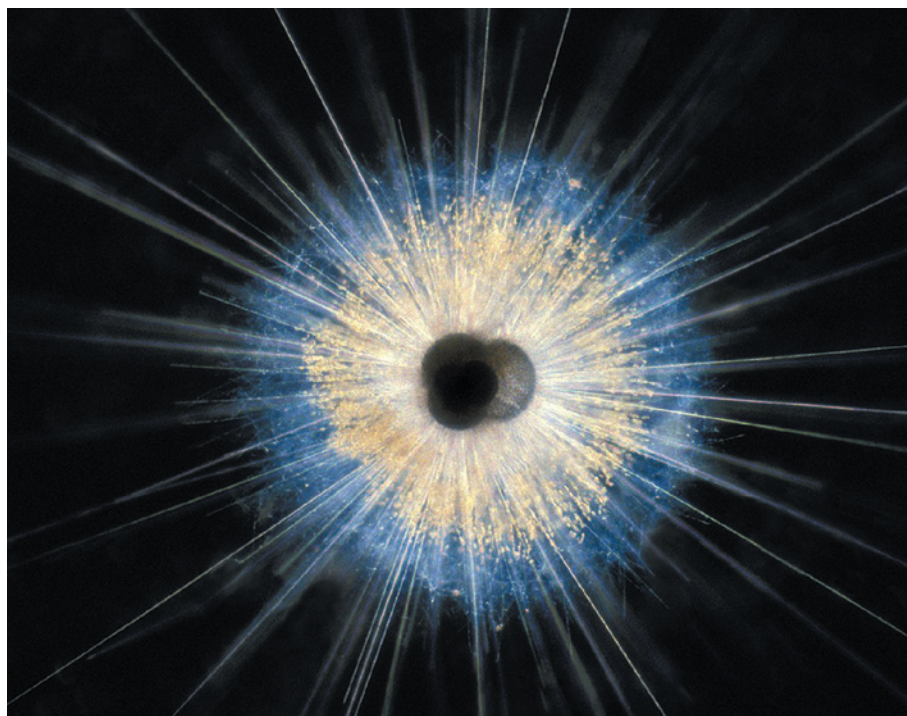
position of cores taken from coral terraces<sup>3</sup>), then fluctuations in global sea level over time can be calculated, as long as ocean sediments provide continuous, undisturbed  $^{18}\text{O}/^{16}\text{O}$  records.

Going back even further in time, a 5-Myr-long composite record<sup>4</sup> of  $^{18}\text{O}/^{16}\text{O}$  ratios from foraminifera that lived in deep-sea sediments was until now considered the best chronicle of ice-sheet volume as Earth shifted from a hot, 'greenhouse' climate (about 55 Myr ago) to colder, 'icehouse' conditions (approximately 2.6 Myr ago). However, there are two problems with this record. First, the proportions of the temperature and ice-volume effects in it are unclear, because deep-ocean temperatures may have changed substantially over this long period of climate transition. Second, it is difficult to scale deep-sea oxygen–isotope variations at orbital timescales into robust estimates (including error margins) for the amplitude of sea-level rise and fall over the past 5 Myr.

Rohling *et al.* overcome these problems by converting  $^{18}\text{O}/^{16}\text{O}$  ratios of fossilized planktonic foraminifera that proliferated in the surface waters of the eastern Mediterranean Sea directly into sea-level variations — an approach previously developed for a study<sup>5</sup> of the Red Sea, and which does not require temperature and ice-volume effects to be disentangled first. Their method depends on a hydraulic model of water exchange through the Strait of Gibraltar, which connects the North Atlantic Ocean and the Mediterranean Sea. This exchange mechanism not only controls the balance of evaporation and water renewal in the Mediterranean, but also strongly affects the seawater oxygen–isotope ratios recorded in planktonic foraminifera.

Assuming that there have been no major tectonic movements in the Strait of Gibraltar during the past 5 Myr that affect its depth and width, the oxygen–isotope signal from these foraminifera is simply a function of global sea-level variations relative to the modern hydraulic state of the Mediterranean Sea. The authors find that estimates of ancient sea levels relative





**Figure 1 | Isotopic recorders.** Planktonic foraminifera, such as *Globigerinoides ruber* (pictured), are unicellular organisms that live near the sea surface. Rohling *et al.*<sup>1</sup> have used oxygen-isotope data from microfossils of planktonic foraminifera from the Mediterranean Sea to deduce global sea levels during the past 5.3 million years.

to modern levels made using their approach are quantitatively similar to those obtained using other approaches, but have much better-defined error statistics. Accordingly, they confirm that sea level was about 10–30 m higher in the warm period that preceded the current ice age than it was when the ice sheets subsequently expanded.

As Rohling and co-workers acknowledge, their method has several limitations that require further study and improvement. Apparent mismatches exist between their results and those of hydraulic models applied to the Red Sea<sup>5</sup>, particularly for periods corresponding to the maximum volume of ice, towards the end of three of the four most recent glacial periods. This could be because the researchers' model for converting planktonic oxygen-isotope data to sea levels might not fully account for freshwater input from rivers, or for perturbations in the oxygen-isotope signal of precipitation and rivers over time. Nonetheless, the striking similarity between Rohling and colleagues' results and those from two independent studies that used data from the Red Sea<sup>5</sup> and the southwest Pacific Ocean<sup>6</sup> to predict sea levels during the past 0.5 and 1.5 Myr, respectively, strongly suggest that the researchers' conclusions for older periods are correct.

As expected for any new approach, the findings leave several open questions. The conclusion that the first major sea-level fall took place more than 0.5 Myr after the onset of significant global cooling 2.75 Myr ago

challenges commonly held assumptions (including some of my own<sup>7</sup>) inferred from the composite record<sup>4</sup> of  $^{18}\text{O}/^{16}\text{O}$  ratios from foraminifera that lived in deep-sea sediments.

#### REPRODUCTIVE BIOLOGY

## Sperm protein finds its mate

Knowledge of the sperm-specific protein that is required for the attachment of sperm to eggs during fertilization in mammals has led to the identification of the protein's receptor on the egg's plasma membrane. [SEE ARTICLE P.483](#)

PAUL M. WASSARMAN

Biology is full of surprises. Such is the case with the findings of Bianchi *et al.*<sup>1</sup> reported on page 483 of this issue. The authors report the end of a decade-long search for a partner for the sperm protein Izumo1, which is responsible for sperm-egg adhesion during fertilization. The elusive mate is a member of the folate-receptor protein family, and is located on the plasma membrane of unfertilized eggs. Because of its essential role in fertilization, the researchers propose that the folate receptor, currently known as Folr4, should be renamed Juno, after the Roman

goddess of fertility and marriage. But it receives a lot of support from other studies of regional palaeoclimatic trends for the period approaching the ice age in the Pleistocene epoch (roughly 2.6 million to 11,700 years ago; see Extended Data Table 2 of the paper<sup>1</sup>).

Rohling *et al.* also conclude that the ice volume during the Last Glacial Maximum — the peak of the most recent glacial period — was larger than that during previous glacial periods. This contradicts continental evidence for the extension of ice sheets over North America and Eurasia<sup>8</sup>, but, if correct, will serve as a crucial benchmark for improved modelling of glacial ice-sheet volumes and heights. Finally, the authors' findings should also lead to improved forecasts for the fate of extant ice caps under global warming. ■

Ralph Schneider is at the Institute of Geosciences, University of Kiel, 24118 Kiel, Germany.

e-mail: [schneider@gpi.uni-kiel.de](mailto:schneider@gpi.uni-kiel.de)

1. Rohling, E. J. *et al. Nature* **508**, 477–482 (2014).
2. Emiliani, C. *Science* **125**, 383–387 (1957).
3. Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
4. Lisiecki, L. E. & Raymo, M. E. *Paleoceanography* **20**, PA1003 (2005).
5. Siddall, M. *et al. Nature* **423**, 853–858 (2003).
6. Elderfield, H. *et al. Science* **337**, 704–709 (2012).
7. Etourneau, J., Schneider, R., Blanz, T. & Martinez, P. *Earth Planet. Sci. Lett.* **297**, 103–110 (2010).
8. Svendsen, J. I. *et al. Quat. Sci. Rev.* **23**, 1229–1271 (2004).

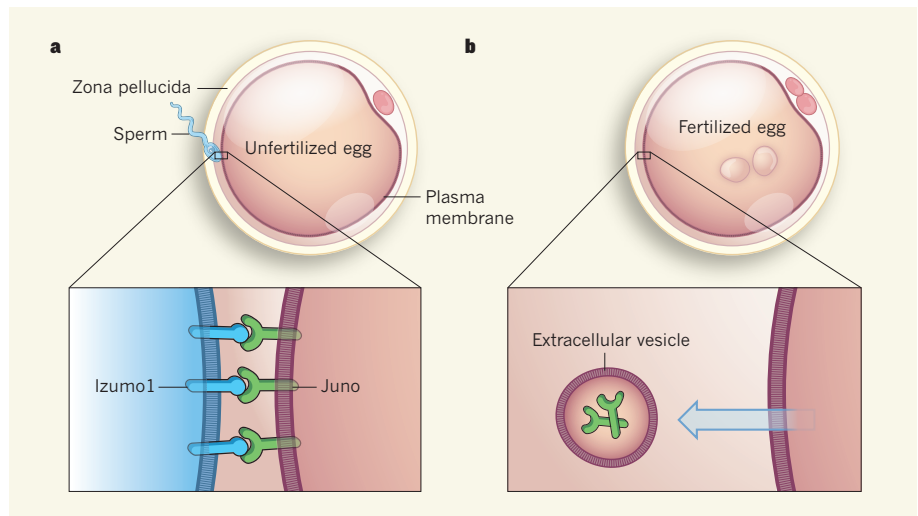
This article was published online on 16 April 2014.

to which Izumo1 binds remained a mystery, owing largely to inherent difficulties in working with small numbers of mammalian eggs, as well as to the transient nature of extracellular interactions between proteins tethered to the membrane.

Bianchi *et al.* identified the mammalian egg receptor for Izumo1 using a method previously developed in their laboratory for detecting weak interactions between membrane-tethered proteins<sup>4</sup>. The authors used a region of the Izumo1 protein — the part that extends into the extracellular space — as a binding probe, and found that a member of the folate-receptor family, Folr4, is the sole receptor for Izumo1 on unfertilized eggs. Furthermore, they confirmed previous suggestions that, unlike other Folr proteins, Folr4 is unable to bind folic acid<sup>5,6</sup> (a form of folate). Bianchi and colleagues found that Folr4, which they rename Juno, is membrane-tethered and is present on the eggs of mice, opossums, pigs and humans.

The researchers examined three Izumo and three folate-receptor family members and found that only Izumo1 and Juno could interact with one another. Exposure of eggs to an antibody that blocks Juno prevented both the binding of Izumo1 to Juno and fertilization *in vitro*. Furthermore, female mice that lacked Juno were infertile, and eggs recovered from these mice could not be fertilized by sperm *in vitro*. In this situation, the authors found that sperm were unable to adhere to and fuse with the egg's membrane, but instead accumulated in the perivitelline space that lies between the egg's membrane and its thick extracellular coat, the zona pellucida. Even when the researchers removed the zona pellucida, exposing the plasma membrane, eggs that lacked Juno could not be fertilized by sperm *in vitro* and fewer sperm bound to the egg's membrane than to the membranes of control eggs. Collectively, these results strongly suggest that the formation of an Izumo1–Juno adhesion complex is essential for fertilization in mammals (Fig. 1a).

These observations beg the question: is the interaction between Izumo1 and Juno the only requirement for sperm–egg fusion? Some recent evidence<sup>7</sup> suggests that, although Izumo1 induces surface interactions such as adhesion between cell membranes, it does not promote membrane fusion. Indeed, when Bianchi and co-workers induced expression of Juno or Izumo1 in non-fusing cell types, they found no evidence for cell fusion. They concluded that the Izumo1–Juno interaction is a necessary event for adhesion between sperm and eggs, but that cell fusion requires other membrane proteins — perhaps similar to those required for fusion in other cell types, such as EFF-1 (ref. 8) and myomaker<sup>9</sup>. The authors go on to propose that local clustering of Juno on the egg's membrane, possibly organized by another membrane-spanning protein such as Cd9 (ref. 10), may occur to increase the strength of sperm binding.



**Figure 1 | An adhesion complex mediates binding of sperm to eggs.** **a**, During fertilization, a single sperm binds to the egg's membrane, which lies under an extracellular coat, the zona pellucida. Bianchi *et al.*<sup>1</sup> report that the protein Izumo1, which is tethered to the membrane of sperm, forms an adhesion complex with its receptor protein, Juno, which spans the egg's membrane. Fertilization does not take place in the absence of this complex. **b**, After fertilization, Juno is lost from the egg's membrane, exiting in extracellular vesicles, thereby preventing the binding and fusion of additional sperm (known as the block to polyspermy).

Bianchi *et al.* got another surprise when they looked for Juno on the fertilized egg's membrane. Thirty to forty minutes after sperm–egg fusion, Juno was barely detectable, having made its exit in extracellular vesicles. The authors suggest that loss of Juno after fertilization may account for the fact that fusion is limited (by an unknown mechanism) to one sperm per egg, a phenomenon known as the block to polyspermy<sup>11</sup> (Fig. 1b). In keeping with this suggestion, Juno did not exit the egg's membrane when sperm were injected directly into eggs (known as intracytoplasmic sperm injection), or when eggs were activated with ethanol. Neither of these treatments results in a block to polyspermy<sup>12</sup> and this could be due to retention of Juno on the fertilized egg's membrane, permitting binding and fusion of additional sperm.

As with other advances in biology, this discovery raises questions and creates possibilities. For example, whether cases of infertility in women can be ascribed specifically to mutations in the gene encoding Juno remains to be evaluated. However, Izumo1 has proved to be a suitable candidate for the development of a contraceptive vaccine<sup>13</sup>, and it is likely that Juno and Izumo1–Juno complexes will provide additional targets for contraception. Detailed knowledge of the three-dimensional structure of the Izumo1–Juno complex could certainly facilitate the design of small molecules that specifically inhibit formation of the adhesion complex and, consequently, prevent fertilization.

Another logical next step is to identify proteins that interact with Izumo1–Juno complexes and stimulate sperm–egg fusion. Further application of the techniques described by Bianchi *et al.* could aid the identification of such proteins. In a similar vein,

the mechanism by which Juno is ejected from the egg membrane after sperm–egg fusion is not yet understood. Removal of Juno in extracellular vesicles may depend on the cortical reaction<sup>14</sup>, a part of the process of preventing polyspermy that involves fusion of vesicles known as cortical granules with the egg's membrane after fertilization. Much remains to be done, but the identification of an egg-based partner for Izumo1 represents a significant step forwards for our understanding of the fertilization process. ■

**Paul M. Wassarman** is in the Department of Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, New York 10029-6574, USA. e-mail: paul.wassarman@mssm.edu

1. Bianchi, E., Doe, B., Goulding, D. & Wright, G. J. *Nature* **508**, 483–487 (2014).
2. Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. *Nature* **434**, 234–238 (2005).
3. Ellerman, D. A. *et al. Mol. Reprod. Dev.* **76**, 1188–1199 (2009).
4. Wright, G. J. *Mol. Biosyst.* **5**, 1405–1412 (2009).
5. Chen, C. *et al. Nature* **500**, 486–489 (2013).
6. Wibowo, A. S. *et al. Proc. Natl Acad. Sci. USA* **110**, 15180–15188 (2013).
7. Inoue, N. *et al. Development* **140**, 3221–3229 (2013).
8. Podbilewicz, B. *et al. Dev. Cell* **11**, 471–481 (2006).
9. Millay, D. P. *et al. Nature* **499**, 301–305 (2013).
10. Jégou, A. *et al. Proc. Natl Acad. Sci. USA* **108**, 10946–10951 (2011).
11. Gardner, A. J. & Evans, J. P. *Reprod. Fertil. Dev.* **18**, 53–61 (2006).
12. Wortzman-Show, G. B., Kurokawa, M., Fissore, R. A. & Evans, J. P. *Mol. Hum. Reprod.* **13**, 557–565 (2007).
13. Wang, M., Lv, Z., Shi, J., Hu, Y. & Xu, C. *Mol. Reprod. Dev.* **76**, 794–801 (2009).
14. Wassarman, P. M. & Litscher, E. S. *Methods Mol. Biol.* **475**, 99–113 (2008).

This article was published online on 16 April 2014.



# Guidelines for investigating causality of sequence variants in human disease

D. G. MacArthur<sup>1,2</sup>, T. A. Manolio<sup>3</sup>, D. P. Dimmock<sup>4</sup>, H. L. Rehm<sup>5,6</sup>, J. Shendure<sup>7</sup>, G. R. Abecasis<sup>8</sup>, D. R. Adams<sup>9,10</sup>, R. B. Altman<sup>11</sup>, S. E. Antonarakis<sup>12,13</sup>, E. A. Ashley<sup>14</sup>, J. C. Barrett<sup>15</sup>, L. G. Biesecker<sup>16</sup>, D. F. Conrad<sup>17</sup>, G. M. Cooper<sup>18</sup>, N. J. Cox<sup>19</sup>, M. J. Daly<sup>1,2</sup>, M. B. Gerstein<sup>20,21</sup>, D. B. Goldstein<sup>22</sup>, J. N. Hirschhorn<sup>2,23</sup>, S. M. Leal<sup>24</sup>, L. A. Pennacchio<sup>25,26</sup>, J. A. Stamatoyannopoulos<sup>27</sup>, S. R. Sunyaev<sup>28,29</sup>, D. Valle<sup>30</sup>, B. F. Voight<sup>31</sup>, W. Winckler<sup>2†</sup> & C. Gunter<sup>18†</sup>

**The discovery of rare genetic variants is accelerating, and clear guidelines for distinguishing disease-causing sequence variants from the many potentially functional variants present in any human genome are urgently needed. Without rigorous standards we risk an acceleration of false-positive reports of causality, which would impede the translation of genomic research findings into the clinical diagnostic setting and hinder biological understanding of disease. Here we discuss the key challenges of assessing sequence variants in human disease, integrating both gene-level and variant-level support for causality. We propose guidelines for summarizing confidence in variant pathogenicity and highlight several areas that require further resource development.**

High-throughput sequencing approaches can generate detailed catalogues of genetic variation in both disease patients and the general population. However, for these technologies to have the greatest medical impact we must be able to separate genuine disease-causing or disease-associated genetic variants reliably from the broader background of variants present in all human genomes that are rare, potentially functional, but not actually pathogenic (Box 1) for the disease or phenotype under investigation.

Many, but unfortunately not all, variants that have been causally associated with rare and common genetic disorders represent robust and correct conclusions. False assignments of pathogenicity can have severe consequences for patients, resulting in incorrect prognostic, therapeutic or reproductive advice, and for the research enterprise, resulting in misallocation of resources for basic and therapeutic research. Unfortunately, although the vast majority of genes reported as causally linked to monogenic diseases are true positives, false assignments of causality at the variant level are a substantial issue. One recent analysis of 406 published severe disease mutations observed in 104 newly sequenced individuals reported that 122 (27%) of these were either common polymorphisms or lacked direct evidence for pathogenicity<sup>1</sup>. Other studies have identified numerous alleged severe-disease-causing variants in the genomes of population controls<sup>2,3</sup>. In other cases, well-powered follow-up studies of high-profile reported mutations have cast serious doubts on initial reports assigning

disease causality to sequence variants<sup>4,5</sup>, but the vast majority of false-positive findings probably remain undetected. As the volume of patient sequencing data increases it is critical that candidate variants are subjected to rigorous evaluation to prevent further misannotation of the pathogenicity of variants in public databases.

This paper describes the challenges in reliably investigating the role of sequence variants in human disease, and approaches to evaluate the evidence supporting variant causality. It represents the conclusions of a working group of experts in genomic research, analysis and clinical diagnostic sequencing convened by the US National Human Genome Research Institute.

We focus on the application of genome-scale approaches to investigating rare germline variants, defined here as variants with a minor allele frequency of <1%. Our recommendations are most relevant for variants with relatively large effects on disease risk. Our intended scope encompasses the vast majority of variants implicated in severe monogenic diseases as well as rare, large-effect risk variants in complex disease<sup>6</sup>, but excludes the common, small-effect variants typically identified by genome-wide association studies of complex traits<sup>7</sup>.

Unambiguous assignment of disease causality for sequence variants is often impossible, particularly for the very low-frequency variants underlying many cases of rare, severe diseases. Consequently, we refer in this manuscript to the concept of implicating a gene or sequence variant: that

<sup>1</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>2</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. <sup>3</sup>Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, Maryland 20892, USA. <sup>4</sup>Division of Genetics, Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA. <sup>5</sup>Laboratory for Molecular Medicine, Partners Healthcare Center for Personalized Genetic Medicine, Cambridge, Massachusetts 02139, USA. <sup>6</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>Department of Genome Sciences, University of Washington, Seattle, Washington 98115, USA. <sup>8</sup>Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>9</sup>NIH Undiagnosed Diseases Program, National Institutes of Health Office of Rare Diseases Research and National Human Genome Research Institute, Bethesda, Maryland 20892, USA. <sup>10</sup>Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>11</sup>Departments of Bioengineering & Genetics, Stanford University, Stanford, California 94305, USA. <sup>12</sup>Department of Genetic Medicine, University of Geneva Medical School, 1211 Geneva, Switzerland. <sup>13</sup>IGE3 Institute of Genetics and Genomics of Geneva, 1211 Geneva, Switzerland. <sup>14</sup>Center for Inherited Cardiovascular Disease, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>15</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1HH, UK. <sup>16</sup>Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA. <sup>17</sup>Departments of Genetics, Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA. <sup>18</sup>HudsonAlpha Institute for Biotechnology, 601 Genome Way, Huntsville, Alabama 35806, USA. <sup>19</sup>Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois 60637, USA. <sup>20</sup>Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut 06520, USA. <sup>21</sup>Departments of Computer Science, Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA. <sup>22</sup>Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA. <sup>23</sup>Divisions of Genetics and Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA. <sup>24</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. <sup>25</sup>Genomics Division, MS 84-171, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>26</sup>US Department of Energy Joint Genome Institute, Walnut Creek, California 94598, USA. <sup>27</sup>Department of Genome Sciences, University of Washington, 1705 Northeast Pacific Street, Seattle, Washington 98195, USA. <sup>28</sup>Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. <sup>29</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>30</sup>McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA. <sup>31</sup>Department of Pharmacology and Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. †Present addresses: Next Generation Diagnostics, Novartis Institutes for BioMedical Research, Cambridge, Massachusetts, USA (W.W.); Marcus Autism Center, Children's Healthcare of Atlanta, Atlanta, Georgia 30329, USA (C.G.).

## BOX 1

## Terms used to describe sequence variants

Lack of clarity in the terms used to describe sequence variants is a major source of confusion in human genetics. We have adopted the following definitions for terms used throughout this manuscript.

**Pathogenic:** contributes mechanistically to disease, but is not necessarily fully penetrant (i.e., may not be sufficient in isolation to cause disease).

**Implicated:** possesses evidence consistent with a pathogenic role, with a defined level of confidence.

**Associated:** significantly enriched in disease cases compared to matched controls.

**Damaging:** alters the normal levels or biochemical function of a gene or gene product.

**Deleterious:** reduces the reproductive fitness of carriers, and would thus be targeted by purifying natural selection.

is, the process of integrating and assessing the evidence supporting a role for that gene or variant in pathogenesis. We emphasize the primacy of strong genetic support for causation for any new gene, which may then be supplemented and extended with ancillary support from functional and informatic studies.

Our recommendations centre on five key areas: study design; gene-level implication; variant-level implication; publication and databases; and implications for clinical diagnosis. Core guidelines for researchers are summarized in Box 2. We also provide a list of factors to consider in the analyses of candidate variants in presumed monogenic diseases (Supplementary Information) and a list of resources for assessing pathogenicity (Supplementary Table 1).

## Study design

Investigators seeking to identify pathogenic variants should select technological and analytical approaches based on the most likely genetic architecture of the disease of interest. Rare, high-penetrance protein-coding variants can be cost-effectively captured by exome sequencing, which is rapidly becoming the first-line approach for presumed monogenic disorders<sup>8</sup>. Cytogenomic arrays and genotyping of linkage panels remain useful approaches for the identification of copy number variation and for identifying co-segregating haplotypes within large Mendelian (especially dominant) disease families, respectively. Optimal approaches to discovering rare pathogenic variants in complex diseases remain unclear: exome sequencing<sup>9</sup>, deep and low-coverage whole-genome sequencing<sup>10</sup> and/or next-generation genotyping arrays with enhanced coverage of protein-coding variants are all being applied in research settings. As the cost of sequencing declines, we expect that deep whole-genome sequencing will soon become the technology of choice for investigating all genetic architectures.

In selecting technological and analytical approaches for a new study, investigators should consider formal power calculations<sup>11</sup> incorporating predicted distributions of allele frequencies and effect sizes for pathogenic variants, genetic and phenotypic heterogeneity of available cohorts, population frequency of the disease, and available sample sizes. Although parameter values may be uncertain, current knowledge of the genetics of the disease and similar traits can be used to constrain likely ranges. In particular, for many diseases there is overwhelming evidence that both locus and allelic heterogeneity is high, such as in autism, epilepsy and schizophrenia. A study design that assumes low locus and allelic heterogeneity would be certain to fail for these conditions, and this fact would be revealed by even casual evaluations of power for reasonable genetic models. Gene discovery for conditions with low locus heterogeneity and sufficiently high-penetrance mutations is occasionally possible by sequencing a single family<sup>12</sup>; however, most gene-discovery applications will require substantially larger sample sizes: multiple unrelated families for rare monogenic

conditions, and thousands to tens of thousands of patients and controls for complex disorders<sup>9,13</sup>.

To assemble large sample sizes will typically require pooling of patient cohorts by multiple investigators. Although such consortium approaches are desirable, investigators should be mindful of systematic differences among cohorts stemming from technical biases, population stratification, and genetic and phenotype heterogeneity. For studies of complex traits, many quality-control methods developed for genome-wide association studies of common variants will also apply to rare variant studies<sup>14</sup>, but DNA sequencing data face a different and typically more challenging set of quality considerations, particularly when data sets are combined for meta-analysis. In addition, new methods may need to be developed to address population stratification of rare variants<sup>15</sup>, which show stronger geographic clustering than common variants<sup>16</sup>; to minimize the impact of stratification, controls should be matched closely to the ancestry of patient samples.

For presumed monogenic diseases, the availability of multiple families with very similar clinical phenotypes substantially increases power for gene discovery. For cases in which there is a single affected proband and no family history, investigators should consider sequencing the unaffected parents of the probands, permitting efficient discovery of *de novo* mutations and compound heterozygous genotypes. Investigators should begin by examining sequence variation in genes known to be associated with that phenotype, and assessing sequence coverage of the coding sequences and splice junctions for these genes before exploring the possibility of new candidate genes in the affected individuals.

## Gene-level implication

To implicate a variant as pathogenic requires that the DNA sequence affected by that variant has a role in the disease process. For genes not previously reported as causal, investigators must simultaneously demonstrate evidence for a role of a candidate gene and one or more variants disrupting it. Even if the candidate gene has been previously implicated in the same or a similar disease phenotype, the overall support from published sources should be carefully assessed and reported. Multiple classes of evidence may potentially contribute to pathogenic inferences at the level of both gene and variant, and include genetic, informatic and experimental data (Table 1 and Supplementary Information). However, in keeping with the history of the field of human genetics, we emphasize the critical primacy of robust statistical genetic support for the implication of new genes, which may then be supplemented with ancillary experimental or informatic evidence supporting a mechanistic role for the gene in the disease in question.

Historically, gene-level implication in monogenic diseases has relied first on identifying a narrow set of candidate genes through genetic data such as linkage analyses or experimental data on biochemical function, and then identifying rare, probably damaging variants (altering the normal levels or biochemical function of a gene or gene product) in one of the candidate genes in multiple affected patients. The increasing availability of large-scale sequencing data now allows genome-scale approaches to gene discovery, in which the distribution of rare, predicted gene-disrupting variants in patients is systematically compared to population controls or well-validated null models to identify genes with an excess of potentially pathogenic variants for clinical and functional follow-up.

It is worth emphasizing that the whole-genome sequence data sets are in some ways more prone to misinterpretation than earlier analyses because of the sheer wealth of candidate causal mutations in any human genome, many of which may provide a compelling story about how the variant may influence the trait; a problem that has been referred to as the 'narrative potential' of human genomes<sup>17</sup>. To avoid such biases the evidence supporting any candidate gene should be contrasted wherever possible with the evidence observed at other presumably non-disease-related genes (for example, by ranking the gene among all others and reporting the probability of a similar or greater contrast being observed by chance). Formal genome-wide statistical approaches to monogenic-disease gene discovery will require considerable methods development, but general



## BOX 2

## Guidelines for implicating sequence variants in human disease

**General guidelines**

- Provide complete positive and negative evidence associated with the gene or variant implication, not just the results that are consistent with pathogenicity.
- In all cases in which it is possible, place genetic, informatic and experimental results within a quantitative framework: determine the probability of observing this result by chance with a randomly selected variant or gene.
- Take advantage of public data sets of genomic variation, functional genomic data and model-organism phenotypes.
- Do not regard prior reports of gene or variant implication as definitive: to the degree that supporting data are available, reassess them as rigorously as your own data.
- Describe and assess clearly the available evidence supporting prior reports of gene or variant implication.

**Assessment of evidence for candidate disease genes**

- In presumed monogenic-disease cases, evaluate genes previously implicated in similar phenotypes before exploring potential new genes.
- Report a new gene as confidently implicated only when variants in the same gene and similar clinical presentations have been confidently implicated in multiple unrelated individuals.
- In all cases in which it is possible, apply statistical methods to compare the distribution of variants in patients with large matched control cohorts or well-calibrated null models.

**Assessment of evidence for candidate pathogenic variants**

- Determine and report the formal statistical evidence for segregation or association of each variant, and its frequency in large control populations matched as closely as possible to patients in terms of ancestry.
- Recognize that strong evidence that a variant is deleterious (in an evolutionary sense) and/or damaging (to gene function) is not sufficient to implicate a variant as playing a causal role in disease.
- Predict variant deleteriousness with comparative genomic approaches, but avoid considering any single method as definitive or multiple methods as independent lines of evidence for implication.
- Validate experimentally the predicted damaging impact of candidate variants using assays of patient-derived tissue or well-established cell or animal models of gene function.
- Avoid assuming that implicated variants are fully penetrant, or completely explanatory in any specific disease case.

**Publications and reporting**

- Assess and report objectively the overall strength and cohesiveness of the evidence supporting pathogenicity for all variants listed in a publication.
- In all cases in which it is possible, ensure that the level of confidence of pathogenicity and supporting evidence are propagated in variant databases.
- Deposit genotype and phenotype data for both controls and disease patients, and for resultant analyses demonstrating associations, in publicly accessible databases, to the maximum degree permissible under study-specific participant consent and ethical approval.
- If returning results for clinical use, highlight strong, actionable findings but also ensure that uncertain or ambiguous findings are clearly conveyed as such, along with appropriate supporting evidence.
- Provide clear cautions regarding decision-making based on variants with limited evidence when the potential for use in medical interventions is high.

guidelines for establishing the significance of variation can be considered here. As we discuss below, these considerations apply equally to assessing the significance of rare variation in common disease studies.

Our paramount recommendation is that for genome-wide analyses of rare variants for both Mendelian and complex disorders, formal calculation of statistical significance should be used to evaluate the strength of evidence of a set of findings, following the well-established standard of maintaining overall type I (false discovery) error rates below 5%. For example, investigators should not simply assume that the presence of two or more independently occurring *de novo* mutations in the same gene within a sequenced cohort is definitive evidence of a causal role for that gene<sup>18,19</sup>; such a threshold results in ever increasing numbers of false positives as the number of sequenced cases increases. To illustrate this, consider the recent situation of four exome sequencing studies, involving a total of 945 families with a child affected by autism<sup>20–23</sup>, which together observed four independent *de novo* missense mutations in *TTN*. Nevertheless, the investigators did not consider *TTN* to have a causal role in autism, and appropriately so: using a statistical model similar to previously published approaches<sup>6,22,24</sup> that accounts for gene size (*TTN* has the largest coding sequence of any gene in the genome), mutation rate, number of trios and distribution of exome coverage, 1.96 *de novo* *TTN* missense or loss-of-function mutations are predicted by chance, which is not significantly different ( $P = 0.14$ ) from the four observed.

We consider a single gene as the fundamental unit for monogenic disease gene testing, for all disease models; a disease caused by *de novo*

mutations or a disease caused by inherited dominant or recessive variants. An appropriate framework for detecting pathogenic variants will evaluate all of the variation in a gene compared to a well-calibrated null model specific for the hypothesis being considered (for example, *de novo*, dominant, recessive).

Although the field has well-established guidelines for declaring significance using linkage data<sup>25</sup>, it is now important to consider a conservative baseline threshold for declaring significance purely from sequencing data of cases, in the absence of other genealogical information. In this scenario, as the gene is the fundamental unit of analysis, and there is no additional data to constrain the search space for genes, a typical study might perform tests on 21,000 protein-coding genes and 9,000 long non-coding RNA genes<sup>26,27</sup>. A conservative genome-wide significance threshold corresponding to this testing strategy is a Bonferroni-corrected  $P$  value of  $1.7 \times 10^{-6}$  (that is, 0.05 out of 30,000). Importantly, if several different schemes are used to define ‘qualifying mutations’ in such analyses, it is necessary to make further statistical adjustments for each of the different sets of rules that are used.

Formal null models can be specified based on the disease model of interest. As mentioned above, the null model for the case of the *de novo* mutation analysis should consider confounding variables such as sample size, gene size and mutation rate (which may vary by orders of magnitude among genes). We note that such null models have power even for extremely rare conditions and small sample sizes: the first exome sequencing study of Kabuki syndrome<sup>28</sup> initially identified 7 *de novo*

**Table 1 | Classes of evidence relevant to the implication of sequence variants in disease**

Evidence level	Evidence class	Examples
<b>Gene level</b>	Genetic	Gene burden: the affected gene shows statistical excess of rare (or <i>de novo</i> ) probably damaging variants segregating in cases compared to control cohorts or null models.
	Experimental	Protein interactions: the gene product interacts with proteins previously implicated (genetically or biochemically) in the disease of interest. Biochemical function: the gene product performs a biochemical function shared with other known genes in the disease of interest, or consistent with the phenotype. Expression: the gene is expressed in tissues relevant to the disease of interest and/or is altered in expression in patients who have the disease. Gene disruption: the gene and/or gene product function is demonstrably altered in patients carrying candidate mutations. Model systems: non-human animal or cell-culture models with a similarly disrupted copy of the affected gene show a phenotype consistent with human disease state. Rescue: the cellular phenotype in patient-derived cells or engineered equivalents can be rescued by addition of the wild-type gene product.
<b>Variant level</b>	Genetic	Association: the variant is significantly enriched in cases compared to controls. Segregation: the variant is co-inherited with disease status within affected families and additional co-segregating pathogenic variants are unlikely or have been excluded. Population frequency: the variant is found at a low frequency, consistent with the proposed inheritance model and disease prevalence, in large population cohorts with similar ancestry to patients.
	Informatic	Conservation: the site of the variant displays evolutionary conservation consistent with deleterious effects of sequence changes at that location. Predicted effect on function: variant is found at the location within the protein predicted to cause functional disruption (for example, enzyme active site, protein-binding region).
	Experimental	Gene disruption: the variant significantly alters levels, splicing or normal biochemical function of the product of the affected gene. This is shown either in patient cells or a well-validated <i>in vitro</i> model system. Phenotype recapitulation: introduction of the variant, or an engineered gene product carrying the variant, into a cell line or animal model results in a phenotype that is consistent with the disease and that is unlikely to arise from disruption of genes selected at random. Rescue: the cellular phenotype in patient-derived cells, model organisms, or engineered equivalents can be rescued by addition of wild-type gene product or specific knockdown of the variant allele.

loss-of-function variants in the *MLL2* gene in just 10 sequenced patients, a finding that is extremely unlikely by chance under the background mutation model described above ( $P = 1.9 \times 10^{-28}$ ) and that provided compelling evidence implicating this gene as causal.

Formal methods for assessing the significance of observations in rare disease cohorts can also be used to assess, for example, the aggregate evidence for segregation of rare variants in a particular gene when considering inherited variation, building on previously published examples<sup>29</sup>. In this case, the null model should be a population genetic model, for instance, the site frequency spectrum (SFS) of variation constructed from a well-matched control cohort. The null model of the SFS for a given gene should consider both the mutation rate and selective constraint acting on that gene. When evaluating data from a single case, the probability that the variation in a gene is from the null model can be estimated by first identifying the most pathogenic class of variant present in that gene in that case, and then by calculating the probability of sampling a variant of the same class of pathogenicity from the null SFS. Similarly, when the recessive disease model applies, the most pathogenic class of variant on the paternal and maternal haplotypes is identified, and then the probability of sampling both variants from the null SFS is calculated. This testing framework for inherited variants is easily scaled to include multiple disease cases. Ideally, to avoid false positives, the control cohort upon which the SFS is based would be sequenced and analysed in a manner identical to the disease cases.

Such methods may not yet be applicable to every rare disease scenario, and will require work to extend to more exotic inheritance modes such as parental imprinting or obligate compound heterozygosity<sup>30</sup>. Although formal methods are established to perform these tests rigorously, researchers should at the very least evaluate and report the level of background variation in an implicated gene in population cohorts, taking advantage of public resources such as the Exome Variant Server (<http://evs.gs.washington.edu/EVS/>) when implicating a new gene in pathogenesis. Furthermore, the analysis of at least some number of controls, sequenced and analysed in a manner identical to cases, can be critical for avoiding the systematic false positives that remain commonplace in exome and genome sequencing.

Just as for genome-wide association studies of common variants<sup>14</sup>, replication of newly implicated disease genes in independent families

or population cohorts is critical supporting evidence, and in most cases essential for a novel gene to be regarded as convincingly implicated in disease. For the rarest disorders additional cases for independent replication may be unavailable and it may be impossible to make a compelling statistical case for implication from human genetic data alone. In these cases, gene implication must be based on an integrated analysis of genetic, informatic and experimental evidence.

Provided that it is carried out in a statistically rigorous fashion, ancillary information can be used to boost power for gene discovery. For example, many genome-wide sequencing-based studies treat all protein-altering variants equally while ignoring all other classes of variants. More elegant schemes aimed at prioritizing based on predicted pathogenicity may boost power for such studies. Another approach is to stratify gene candidates by their expression in a tissue appropriate to the disease under analysis. For example, a recent study combined variant- and gene-level stratification to show that the *de novo* mutation rate in congenital heart disease was similar in cases versus controls, but the odds ratio rose to 7.5 when focusing on *de novo* mutations predicted to be damaging and to occur in genes expressed in the developing heart<sup>31</sup>.

Experimental evidence that can contribute to support for gene implication falls into three broad categories, listed here in order of increasing strength. First, experimental data can be used to demonstrate that the normal function of the gene is consistent with the known biology of the disease process, for example by showing that the gene is expressed in tissues relevant to the disease<sup>32</sup>, or that its protein product co-localizes with, or physically interacts with, the products of other genes previously implicated in the disease<sup>33</sup>. Second, investigators can demonstrate that a gene product is functionally disrupted by mutations in patients with the disease of interest, as discussed in the variant-level evidence section below. Lastly, disruption of the candidate gene in a model organism can be shown to result in a phenotype that recapitulates the relevant pathology in humans and is unlikely to occur with disruption of genes selected at random<sup>34,35</sup>.

A complete description of the experimental methods relevant to gene implication falls outside the scope of this manuscript. However, we note that the value of experimental approaches depends critically on the appropriateness of the model system to the human disorder that is being investigated. Whether cell line or animal models will be most appropriate will



depend on context: simple cultured cell models may be inappropriate for developmental disorders affecting complex organ systems. For similar reasons, animal models are not well suited for analysis of human-specific aspects of biology.

As noted above, it is also important to consider the specificity of gene-level support; that is, the probability of observing a similar result if the experiment or analysis was performed with a randomly selected gene. For example, if a new candidate gene is implicated in non-syndromic short stature in humans, observing that its orthologue is associated with small body size in knockout mice is relatively uninformative given that a similar phenotype occurs in over 30% of all knockout mouse strains<sup>36</sup>. Similarly, reports that the product of a gene potentially implicated in a metabolic disorder is localized to mitochondria should also consider that these are complex organelles with many highly expressed genes. Wherever possible, investigators should use informatics approaches to assess such metrics in publicly available high-throughput data sets of functional genomic and model organism phenotype data<sup>37</sup>. Although it remains challenging to quantify the statistical confidence of functional observations, those that can be convincingly demonstrated to represent very low-probability events under an appropriate null hypothesis provide more compelling support for implicating a given variant. Even in situations in which a formal statistical framework is not possible we emphasize that researchers must assess functional data rigorously and clearly report their limitations.

### Variant-level implication

Genetic evidence implicating a variant must be assessed within the context of the considerable background of rare genetic variants in humans. Even healthy individuals carry many rare protein-disrupting variants<sup>38</sup>, and about half carry at least one *de novo* protein-altering mutation<sup>39</sup>. Such variants are therefore not typically sufficient proof of causality when observed in a disease case, even if present in well-established disease genes: genes differ markedly in their tolerance to variation<sup>40</sup> and rare variants predicted to be damaging in disease-associated genes are often observed even in population controls<sup>41</sup>.

In both established and newly implicated disease genes, investigators should formally assess and report the statistical support for association. Family-based studies should also assess co-segregation of candidate variants with disease status. Given that a separate, unobserved pathogenic mutation may lie on the same haplotype as the candidate variant, segregation analysis alone cannot definitively implicate a specific variant as pathogenic, but (at least under an assumption of complete penetrance) lack of segregation can exclude non-pathogenic variants from consideration.

Informatic and/or experimental evidence for variant implication can be used to assess whether a variant is likely to be deleterious in an evolutionary sense (Box 1), which primarily comes from *in silico* annotation and comparative genomics<sup>42</sup>, and predict that a variant is damaging in terms of biological function, arising both from computational predictions and experimental assays. Both categories of evidence can support implication, but they do not necessarily demonstrate a causal role for the variant with respect to the trait under study. Again, we stress that hundreds to thousands of coding variants in an individual will typically be labelled as potentially deleterious or damaging, or both; the strength of the resulting evidence for pathogenicity must be considered in the context of this background level of variation.

Measures of evolutionary sequence conservation are widely used indicators of deleteriousness for both protein-coding and non-coding variation<sup>42</sup>. Such approaches have demonstrated value in prioritizing candidate variants<sup>43,44</sup>; however, their predictive power is limited by both statistical and biological factors. Many deleterious variants do not show a strong conservation signature, particularly if the gene has been subject to rapid evolution in the human or primate lineage, or if there have been compensatory substitutions in other regions of the protein in ancestral species<sup>45</sup>. Conversely, strong conservation can be maintained at sites subject to even relatively weak selective pressure, at which variants may have only small effects on disease risk. The power of these methods also depends on the accuracy and phylogenetic scope of the underlying sequence alignments.

These limitations should be taken into account when using predictions of deleteriousness as evidence for implication. Even though it is worthwhile to use multiple prediction algorithms, investigators should avoid treating these as though they represent strong or independent lines of evidence for pathogenicity.

Although some classes of variation, such as truncating or splice-site-disrupting variants in the middle of a protein-coding gene, are more likely to be damaging than others, such variants are also enriched for sequencing and annotation errors and may be rescued by alternative RNA splicing, other variants, or local sequence context<sup>41</sup>. These possibilities should be assessed, and if possible the predicted damaging effect should be confirmed experimentally.

Experimental approaches to investigating the impact of a sequence variant on gene function, or cell or organism phenotype, can also have a role in demonstrating that a variant is damaging to gene function and in identifying the molecular mechanisms underlying a variant's effect on disease risk. However, great care must be taken to select appropriate experimental methods, which will depend on the class of variant, biological context (for example, tissue type), access to samples and reagents, desired throughput, time and cost. When a gene has already been confidently implicated in disease, and it is known what class of variant is causal (for instance, loss or gain of function as represented by a specific assay), then an experiment that places a variant of unknown significance into such a functional class can be particularly informative.

Evidence derived directly from patient tissue or cells can often be stronger than that from model systems, particularly (for loss-of-function variants) if the molecular defect can be rescued by complementation in a cellular assay. Replicating disease-relevant phenotypes in a heterologous cell line engineered to carry the proposed causal variant can help to rule out effects of a patient's genetic background on disease outcome. Weaker but still valuable support can be provided by assays performed in model organisms, more artificial cell culture systems, and non-cellular models such as construct-based assays of altered protein–protein interactions or transcript splicing. Models are most valuable if they directly mimic the predicted functional impact of the candidate variants: for example, knockout mice are better models of recessive loss of function than of dominant missense mutations in a candidate gene. In the case of compound heterozygous recessive inheritance—particularly if the proposed mode of action depends on an interaction between allelic variants, such as in TAR (thrombocytopenia with absent radius) syndrome<sup>30</sup>—it will be necessary to develop cellular assays that incorporate and assess multiple variants simultaneously.

The impact of variation in non-protein-coding regions of the genome—such as splicing and transcriptional enhancers—remains particularly challenging to interpret, but we note that systematic experimental approaches have begun to both highlight the regions of the human genome most likely to have a role in gene regulation<sup>46</sup>, and to dissect the potential impact of variation within them<sup>47</sup>. However, given the challenges of predicting impact for non-coding variants, it remains critical to determine whether the purported pathogenic variant does in fact produce the expected effect on expression or splicing of the affected gene, either by demonstrating an unusual expression level in the patient or by *in vitro* experimentation (such as minigene constructs).

We caution against the assumption that convincingly implicated variants, even in presumed monogenic disorders, are necessarily fully penetrant (that is, sufficient in isolation to cause disease). In fact the penetrance of most reported disease-associated mutations has not been accurately assessed with current data owing to the biases associated with sample ascertainment. Indeed, the prevalence of reported severe-disease-causing mutations in population controls<sup>2,3</sup> suggests that incomplete penetrance, false assignment of pathogenicity, or wider-than-appreciated ranges of expressivity are a substantially more common feature of reported Mendelian disease mutations than generally appreciated. Accurate estimates of penetrance require characterization of reported mutations in large, well-phenotyped population cohorts<sup>48–50</sup>. Further large-scale studies of this kind should be a priority for the field.

We also note the underappreciated importance of calibrating the accuracy of functional assays by large-scale testing of variants confidently established to be non-pathogenic (for example, common missense polymorphisms in the gene of interest). Such experiments establish a baseline estimate for the impact of well-tolerated variants on the assay in question.

## Publication and data sharing

As noted above, there are many false positives in disease-mutation databases, stemming largely from erroneous assignment of pathogenicity both in clinical diagnostic laboratories and in the primary literature<sup>1,2,51</sup>. To reduce this burden will require robust, centralized repositories of mutation data, incorporating explicit, structured evidence for variant pathogenicity and systems for rapid correction of entries. To incentivize both research and clinical laboratories to deposit variation data into open repositories, and to update evidence for or against implication, is a key challenge to be addressed by funding bodies, journals, research consortia, clinical organizations and others<sup>52</sup>. We are hopeful that such activities can be coordinated around the US National Center for Biotechnology Information (NCBI)'s newly launched ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>), which will also interface with existing efforts in this space including the LOVD (Leiden Open (source) Variation Database)<sup>53</sup> and other locus-specific databases, OMIM (Online Mendelian Inheritance in Man; <http://omim.org/>) and DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources)<sup>54</sup>.

In some cases—such as diseases that are extremely rare or have high degrees of locus heterogeneity—it may be impossible to obtain definitive evidence implicating a specific gene or variant with available sample sizes. In such cases we acknowledge that the suggestive evidence pointing to a gene's potential implication can nevertheless be valuable in future clinical and research investigations, and should not be excluded from publications or the public domain. However, it is incumbent on investigators, reviewers and journals to be explicit in describing the supporting evidence and the degree of confidence in causality for each proposed gene association and reported variant.

Finally, we emphasize the value of sharing sequence and phenotype data from clinical and research samples to the fullest possible extent. Many investigators and research funders consider responsible data sharing to be a moral and professional imperative<sup>55</sup>. In many cases, particularly for extremely rare phenotypes, individual laboratories that are not actively recruiting subjects will evaluate only a handful of samples. Sharing of sequence data among testing laboratories has often been restricted, so that many potentially pathogenic mutations and associated phenotypes are known only to individual laboratories. The availability of genome-wide variant calls and detailed clinical phenotype descriptions from such patients in centralized repositories—which will require substantial investment both in informatic infrastructure and new ethical frameworks—would permit more rapid accumulation of evidence for novel genes, and continuous reanalysis to refine the classification of potentially implicated variants and the genotype–phenotype map of human disease. Models for successful data sharing efforts in rare disease already exist in the field of copy number variation with the DECIPHER database<sup>54</sup> and the International Standards for Cytogenomic Arrays Consortium (<https://www.iscaconsortium.org/>), aided by an increasing number of rare-disease resource consortia, and several ambitious efforts to establish clear global standards for genomic data sharing are now underway<sup>56</sup>.

## Added challenges in clinical settings

Although this summary is focused on research, research findings provide the foundation for clinical interpretation. Questionable attributions of causality based on weak research evidence can be readily propagated through research databases and can be misinterpreted clinically as stronger than they truly are. Thus, even researchers who do not explicitly provide diagnosis to patients should be aware that their published findings may be used as support for decisions made in clinical settings.

Clinical laboratories face similar challenges in assessing variant pathogenicity as do researchers, but with the added pressures of diagnostic

urgency and the potentially severe consequences of misdiagnosis. Although guidelines are available for variant interpretation in a diagnostic setting<sup>57</sup>, analytical frameworks for next-generation sequencing data are only beginning to emerge<sup>58,59</sup>. Responsible application of these technologies will require standards for test validation, variant interpretation and return of results.

The results of genetic and genomic testing are increasingly being used in medical decision-making, including recommendations for prophylactic mastectomy, cardiac defibrillator implantation, tumour therapy and prenatal diagnosis. These actions are neither generally inappropriate nor uniformly incorrect; however, the potential for harm due to misinterpretation of variants is substantial. Although physicians must often make medical decisions using imperfect or ambiguous data, it is critical that healthcare providers be made aware of the varying levels of certainty in the evidence for implicating a variant in disease, both through the consistent use of variant classification terminologies and descriptions of the supporting evidence or lack thereof.

## Conclusions

High-throughput DNA sequencing technologies provide unprecedented opportunities to discover new genes and variants underlying human disease, but these discoveries must be rigorously performed and replicated to prevent the proliferation of false-positive findings.

Assessment of evidence for variant implication is a two-step process. First, the overall evidence for implication of a gene should be considered, focusing primarily on the statistical support for implication from genetic analyses, potentially supplemented by ancillary data from informatic sources and functional studies. Second, a combined assessment of the genetic, experimental and informatic support for individual candidate variants should be performed. Such assessments should be performed even if the genes or variants have been previously reported as confidently implicated; prior evidence should be continuously re-evaluated with newly available information.

We urge that, whenever possible, investigators assess the results of genetic, informatic and functional analyses within a quantitative statistical framework, such as determining the probability of the observed distribution of genetic variants in cases and controls under the null hypothesis, and the a priori power to detect variants of a specified frequency and effect size. The specificity of experimental or informatic results provided in support of implication should also be assessed whenever possible by asking how often a similar result would be obtained by chance among a set of random variants or genes. In such analyses investigators should

### BOX 3

## Priorities for research and infrastructure development

- Improved public databases of human genetic variants incorporating explicit, up-to-date supporting evidence for variant implication in disease and audit trails recording changes in interpretation.
- Improved incentives, and ethical and logistical solutions, for sharing of genetic and phenotypic data from both research and clinical diagnostic laboratories.
- Public databases of variant and allele frequency data from large sets of population reference samples from a wide range of ancestries.
- Large-scale genotyping of reported human disease-causing variants in large, well-phenotyped population cohorts, reducing biases in the assessment of the associated penetrance and phenotypic heterogeneity.
- Development and benchmarking of standardized, quantitative statistical approaches for objectively assigning probability of causation to new candidate disease genes and variants.



take advantage of the increasing availability of genome-scale sequencing and functional data, and help to build these resources by contributing their findings to public databases.

The community should also focus on the ongoing development of resources in several key areas (Box 3). In particular, major improvements in databases of reported pathogenic mutations, including details of the evidence supporting pathogenicity, are urgently needed. Large-scale experiments to assay previously reported disease-associated mutations in additional large, well-phenotyped populations will also be required to confirm pathogenicity and provide robust evidence of penetrance and expressivity. Finally, extensive work is needed to develop formal statistical frameworks for quantifying the strength of the evidence for implication.

Objective, systematic and quantitative evaluation of the evidence for pathogenicity and sharing of these evaluations and data amongst research and clinical laboratories will maximize the chances that disease-causing genetic variants are correctly differentiated from the many rare non-pathogenic variants seen in all human genomes.

Received 24 June 2013; accepted 5 February 2014.

1. Bell, C. J. *et al.* Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci. Transl. Med.* **3**, 65ra4 (2011).
  2. Xue, Y. *et al.* Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.* **91**, 1022–1032 (2012).
  3. Norton, N. *et al.* Evaluating pathogenicity of rare variants from dilated cardiomyopathy in the exome era. *Circ. Cardiovasc. Genet.* **5**, 167–174 (2012).
  4. Weng, L. *et al.* Lack of MEF2A mutations in coronary artery disease. *J. Clin. Invest.* **115**, 1016–1020 (2005).
  5. Hunt, K. A. *et al.* Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. *Nature Genet.* **44**, 3–5 (2012).
  6. Allen, A. S. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
  7. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
  8. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Rev. Genet.* **12**, 745–755 (2011).
  9. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genet.* **44**, 623–630 (2012).
  10. Pasaniuc, B. *et al.* Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genet.* **44**, 631–635 (2012).
  11. Li, B., Wang, G. & Leal, S. M. SimRare: a program to generate and analyze sequence-based data for association studies of quantitative and qualitative traits. *Bioinformatics* **28**, 2703–2704 (2012).
  12. Johnston, J. J. *et al.* The phenotype of a germline mutation in PIGA: the gene somatically mutated in paroxysmal nocturnal hemoglobinuria. *Am. J. Hum. Genet.* **90**, 295–300 (2012).
  13. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111**, E445–E464 (2014).
  14. Chanock, S. J. *et al.* Replicating genotype–phenotype associations. *Nature* **447**, 655–660 (2007).
  15. O'Connor, T. D. *et al.* Fine-scale patterns of population stratification confound rare variant association tests. *PLoS ONE* **8**, e65834 (2013).
  16. Mathieson, I. & McVean, G. Differential confounding of rare and common variants in spatially structured populations. *Nature Genet.* **44**, 243–246 (2012).
  17. Goldstein, D. B. *et al.* Sequencing studies in human genetics: design and interpretation. *Nature Rev. Genet.* **14**, 460–470 (2013).
  18. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
  19. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
  20. Sanders, S. J. *et al.* De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
  21. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485**, 246–250 (2012).
  22. Neale, B. M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
  23. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
  24. O'Roak, B. J. *et al.* Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **1619–1622** (2012).
  25. Lander, E. & Kruglyak, L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
  26. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
  27. Derrien, T. *et al.* The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* **22**, 1775–1789 (2012).
  28. Ng, S. B. *et al.* Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nature Genet.* **42**, 790–793 (2010).
  29. Lemaire, M. *et al.* Recessive mutations in *DGKE* cause atypical hemolytic-uremic syndrome. *Nature Genet.* **45**, 531–536 (2013).
- This paper provides useful statistical approaches to assess the strength of the genetic evidence for a recessive disease gene.**
30. Albers, C. A. *et al.* Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon–junction complex subunit *RBM8A* causes TAR syndrome. *Nature Genet.* **44**, 435–439 (2012).
  31. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
  32. Lage, K. *et al.* A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes. *Proc. Natl Acad. Sci. USA* **105**, 20870–20875 (2008).
  33. Franke, L. *et al.* Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. *Am. J. Hum. Genet.* **78**, 1011–1025 (2006).
  34. Boulding, H. & Webber, C. Large-scale objective association of mouse phenotypes with human symptoms through structural variation identified in patients with developmental disorders. *Hum. Mutat.* **33**, 874–883 (2012).
  35. Webber, C. *et al.* Forging links between human mental retardation-associated CNVs and mouse gene knockout models. *PLoS Genet.* **5**, e1000531 (2009).
  36. Reed, D. R., Lawler, M. P. & Tordoff, M. G. Reduced body weight is a common effect of gene knockout in mice. *BMC Genet.* **9**, 4 (2008).
  37. Giallourakis, C., Henson, C., Reich, M., Xie, X. & Mootha, V. K. Disease gene discovery through integrative genomics. *Annu. Rev. Genomics Hum. Genet.* **6**, 381–406 (2005).
  38. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
  39. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nature Rev. Genet.* **13**, 565–575 (2012).
  40. Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human genome. *Nature* **437**, 1153–1157 (2005).
  41. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
  42. Cooper, G. M. & Shendure, J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Rev. Genet.* **12**, 628–640 (2011).
  43. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
  44. Cooper, G. M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nature Methods* **7**, 250–251 (2010).
  45. Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky–Muller incompatibilities in protein evolution. *Proc. Natl Acad. Sci. USA* **99**, 14878–14883 (2002).
  46. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
  47. Patwardhan, R. P. *et al.* Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnol.* **30**, 265–270 (2012).
  48. Cooper, G. M. *et al.* A copy number variation morbidity map of developmental delay. *Nature Genet.* **43**, 838–846 (2011).
  49. Bick, A. G. *et al.* Burden of rare sarcomere gene variants in the Framingham and Jackson Heart Study cohorts. *Am. J. Hum. Genet.* **91**, 513–519 (2012).
  50. Flannick, J. *et al.* Assessing the phenotypic effects in the general population of rare variants in genes for a dominant Mendelian form of diabetes. *Nature Genet.* **1380–1385** (2013).
- One of the first papers to explore systematically the impact of normal human genetic variation in Mendelian disease genes; the paper shows that many previously reported severe disease mutations are not in fact completely penetrant.**
51. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010); corrigendum, **473**, 544 (2011).
  52. Editorial. Share alike. *Nature* **490**, 143–144 (2012).
  53. Fokkema, I. F. *et al.* LOVD v2.0: the next generation in gene variant databases. *Hum. Mutat.* **32**, 557–563 (2011).
  54. Firth, H. V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
  55. Walport, M. & Brest, P. Sharing research data to improve public health. *Lancet* **377**, 537–539 (2011).
  56. Global Alliance for Genomics and Health. Creating a global alliance to enable responsible sharing of genomic and clinical data. <http://genomicsandhealth.org/files/public/White%20Paper%20June%202013%20final.pdf> (2013).
  57. Richards, C. S. *et al.* ACMG recommendations for standards for interpretation and reporting of sequence variations: revisions 2007. *Genet. Med.* **10**, 294–300 (2008).
  58. Gargis, A. S. *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nature Biotechnol.* **30**, 1033–1036 (2012).
  59. Rehms, H. L. *et al.* ACMG Clinical Laboratory Standards for Next Generation Sequencing. *Genet. Med.* (in the press) (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This paper was inspired by the deliberations of an expert working group convened by the US National Human Genome Research Institute (NHGRI) on 12 and 13 September 2012 to address the challenges of assigning disease causality to genetic variants. The authors acknowledge B. M. Neale, L. E. Duncan, K. E. Samocha, E. T. Lim and C. G. MacArthur for contributions to the manuscript.

**Author Contributions** D.G.M., T.A.M. and C.G. planned the project and led the writing group. D.G.M., T.A.M., C.G., D.P.D., H.L.R. and J.S. served as the organizing

committee. D.G.M., T.A.M., D.P.D., H.L.R., J.S., G.R.A., D.R.A., R.B.A., S.E.A., E.A.A., J.C.B., L.G.B., D.F.C., G.M.C., N.J.C., M.J.D., M.B.G., D.B.G., J.N.H., S.M.L., L.A.P., J.A.S., S.R.S., D.V., B.F.V., W.W. and C.G. attended a September 2012 workshop, contributed guidelines from their own expertise, and reviewed and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests.

Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.G.M. ([macarthur@atgu.mgh.harvard.edu](mailto:macarthur@atgu.mgh.harvard.edu)) or C.G. ([drchrisgunter@gmail.com](mailto:drchrisgunter@gmail.com)).



This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>



# Sea-level and deep-sea-temperature variability over the past 5.3 million years

E. J. Rohling<sup>1,2</sup>, G. L. Foster<sup>2</sup>, K. M. Grant<sup>1</sup>, G. Marino<sup>1</sup>, A. P. Roberts<sup>1</sup>, M. E. Tamisiea<sup>3</sup> & F. Williams<sup>2</sup>

**Ice volume (and hence sea level) and deep-sea temperature are key measures of global climate change. Sea level has been documented using several independent methods over the past 0.5 million years (Myr). Older periods, however, lack such independent validation; all existing records are related to deep-sea oxygen isotope ( $\delta^{18}\text{O}$ ) data that are influenced by processes unrelated to sea level. For deep-sea temperature, only one continuous high-resolution (Mg/Ca-based) record exists, with related sea-level estimates, spanning the past 1.5 Myr. Here we present a novel sea-level reconstruction, with associated estimates of deep-sea temperature, which independently validates the previous 0–1.5 Myr reconstruction and extends it back to 5.3 Myr ago. We find that deep-sea temperature and sea level generally decreased through time, but distinctly out of synchrony, which is remarkable given the importance of ice–albedo feedbacks on the radiative forcing of climate. In particular, we observe a large temporal offset during the onset of Plio–Pleistocene ice ages, between a marked cooling step at 2.73 Myr ago and the first major glaciation at 2.15 Myr ago. Last, we tentatively infer that ice sheets may have grown largest during glacialials with more modest reductions in deep-sea temperature.**

To understand better the potential response of ice volume (sea level) to global warming, there is a need for continuous, highly resolved and well-quantified records of sea-level variations associated with past climate fluctuations<sup>1–4</sup>. In addition, such records are critical for understanding the development of major ice-age cycles over the past  $\sim 3$  Myr and of the attendant reorganizations in the coupled climate–ocean system, including extensive biological and biogeochemical perturbations<sup>5–8</sup>.

Continuous sea-level records with centennial resolution, suitable for investigating magnitudes and rates of sea-level change, exist for the past 0.5 Myr (refs 1, 9–11). For older periods, existing millennially resolved sea-level reconstructions include (1) a continuous, model-based deconvolution of global deep-sea benthic foraminiferal  $\delta^{18}\text{O}$  data ( $\delta^{18}\text{O}_b$ ) into temperature and ice-volume changes back to 35 Myr ago<sup>12</sup>; (2) a scaling of global deep-sea  $\delta^{18}\text{O}_b$  using New Zealand sequence stratigraphic data for 3.4–2.3 Myr ago<sup>2,3</sup>; (3) a direct scaling of another deep-sea  $\delta^{18}\text{O}_b$  compilation for the past 7 Myr (ref. 13); and (4) a deep-sea  $\delta^{18}\text{O}_b$  record (corrected for deep-sea temperature,  $T_{ds}$ ) over the past 1.5 Myr from Chatham rise, in the southwest Pacific Ocean, which is argued to be representative of global deep water<sup>14</sup>. A further, lower-resolution,  $T_{ds}$ -corrected deep-sea  $\delta^{18}\text{O}_b$  record exists for the North Atlantic<sup>15</sup>, although questions exist concerning carbonate chemistry influences on the epibenthic species analysed in that record<sup>16</sup>, and about site-specific issues regarding North Atlantic Deep Water property variations versus potential water-mass changes due to Antarctic Bottom Water penetration.

All existing methods rely on deep-sea  $\delta^{18}\text{O}_b$  and are, therefore, not independent of each other. In addition, there is limited temporal overlap between these reconstructions and, importantly, methodological uncertainties are typically much larger in sea-level reconstructions for periods before 0.5 Myr ago. Hence, it is necessary to develop independent sea-level reconstructions to identify mutually consistent patterns. This in turn will enable fundamental questions to be addressed concerning the timing and development of Northern Hemisphere glaciation, sea-level variability during past warm periods with greenhouse gas concentrations similar to those of today, and the long-term relationship between ice volume, temperatures and greenhouse gas concentrations.

To advance the debate, we here present a new and independent sea-level reconstruction that spans the past 5.3 Myr.

## Location for new sea-level reconstruction

The Red Sea would be a promising location for developing an extended sea-level reconstruction beyond 0.5 Myr ago<sup>1,10,11,17</sup>. Red Sea sea-level reconstructions for the past 0.5 Myr rely on hydraulic control of water exchange through a shallow and narrow connection with the open ocean (the Bab-el-Mandab Strait)<sup>10,17</sup>. The method is independent of deep-sea  $\delta^{18}\text{O}_b$ , and yields ‘Relative sea level at Bab-el-Mandab’ ( $\text{RSL}_{\text{BeM}}$ ) reconstructions with a  $1\sigma$  uncertainty of  $\sim 6$  m (refs 10, 17). Unfortunately, no high-quality Red Sea sediment cores exist that allow extension of  $\text{RSL}_{\text{BeM}}$  beyond 0.55 Myr ago. We therefore shift focus to the Mediterranean Sea, which is another evaporative marginal sea with limited connection to the open ocean.

Discerning a sea-level signal in Mediterranean records of carbonate microfossil  $\delta^{18}\text{O}$  is more complex than in the Red Sea because (1) the larger strait profile at Gibraltar, relative to Bab-el-Mandab, causes a lower signal-to-noise ratio, with Mediterranean glacial–interglacial  $\delta^{18}\text{O}$  amplitudes of 2.5–3‰ compared to Red Sea amplitudes of 5.5–6‰; and (2) the Mediterranean hydrological cycle is more complicated than in the Red Sea, with major rivers that integrate information from a large catchment area with influences from both temperate climate conditions and the African monsoon<sup>10,11,17–19</sup>. However, the Mediterranean provides uninterrupted sediment records dating back to the end of the Messinian salinity crisis at 5.33 Myr ago, from tectonically uplifted marine sediments and long deep-sea sediment cores<sup>20,21</sup>. Eastern Mediterranean planktonic foraminiferal  $\delta^{18}\text{O}$  records ( $\delta^{18}\text{O}_p$ ) have recently been synthesized into a 5.3-Myr ‘Mediterranean stack’ with a millennially resolved, orbitally tuned chronology<sup>22</sup>. It is particularly beneficial for sea-level reconstruction that orbital tuning of the Mediterranean record (in contrast to deep-sea  $\delta^{18}\text{O}_b$  records) makes no assumptions about the relationship between insolation and ice volume; instead, it employs a timing relationship between insolation and African monsoon intensity (using sedimentary cycles)<sup>20</sup>. The Mediterranean chronology is so

<sup>1</sup>Research School of Earth Sciences, The Australian National University, Canberra 0200, Australia. <sup>2</sup>Ocean and Earth Science, University of Southampton, National Oceanography Centre, Southampton SO14 3ZH, UK. <sup>3</sup>National Oceanography Centre, Joseph Proudman Building, Liverpool L3 5DA, UK.

well established that it underpins global geochronology throughout the time interval considered here<sup>20,23</sup>. Therefore, the eastern Mediterranean  $\delta^{18}\text{O}_p$  stack<sup>22</sup> is an excellent resource for developing a long sea-level record using a Mediterranean version of the method that was developed for the Red Sea (see Methods). This has only recently become possible owing to increased quantitative understanding of the relationship between hydrological processes and  $\delta^{18}\text{O}$  changes in and around the Mediterranean<sup>11,18,19,24</sup>.

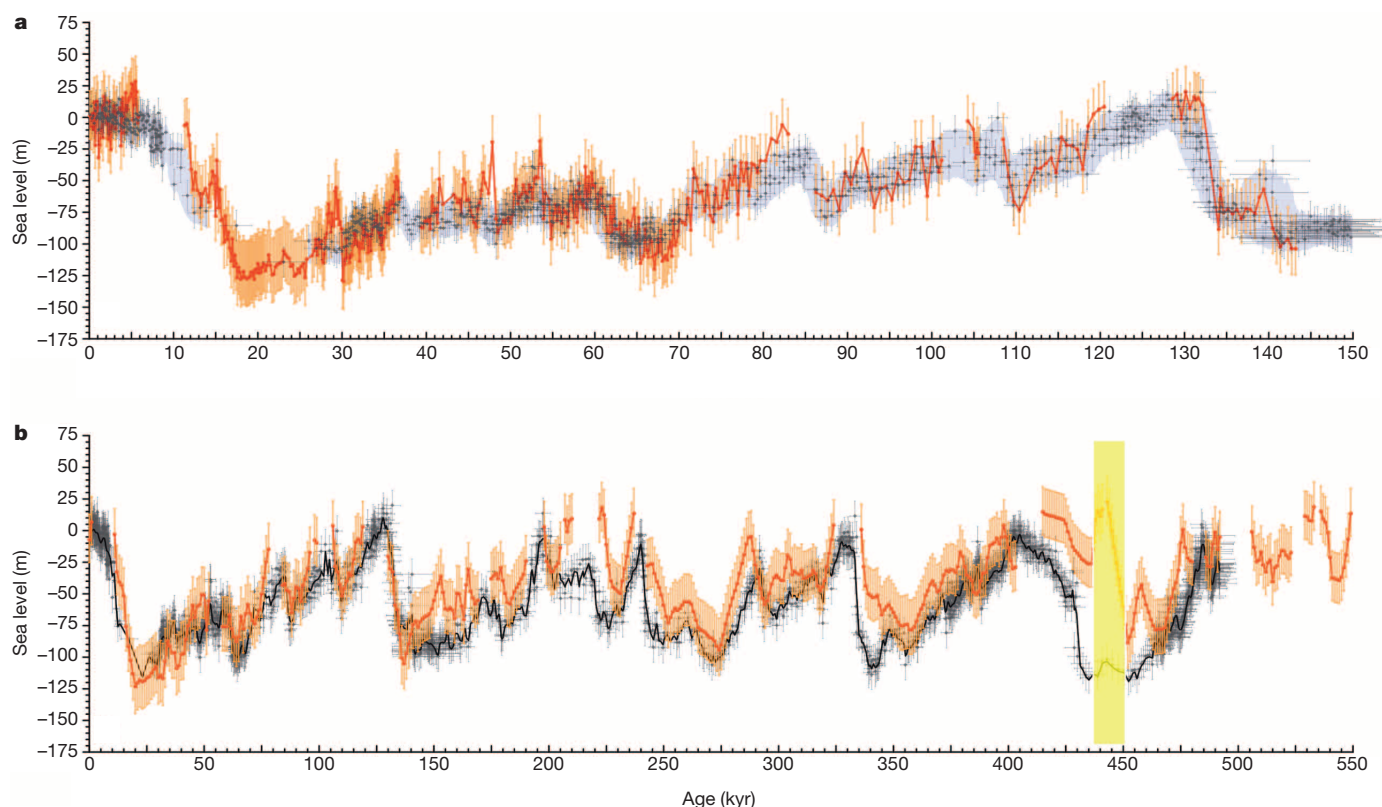
### Converting Mediterranean $\delta^{18}\text{O}$ to sea level

We quantify relative sea-level changes at Gibraltar (RSL<sub>Gib</sub>) using values of eastern Mediterranean  $\delta^{18}\text{O}_p$  after removal of ‘sapropel’ intervals of major surface freshwater dilution. These intervals are associated with periods of sea-floor anoxia and are typically marked by dark olive to black organic-rich sediments (bounded by pale organic-poor deposits), with light surface-water  $\delta^{18}\text{O}$  anomalies, elevated sedimentary Ba/Al ratios and an absence of benthic microfossils<sup>19,21,25–29</sup>. We exclude sapropel intervals from RSL<sub>Gib</sub> on the basis of a combination of  $\delta^{18}\text{O}_p$  anomaly detection and visual evidence (Extended Data Fig. 1).

Our RSL<sub>Gib</sub> reconstruction method is explained in detail in Methods (with code in Supplementary Information). The method exploits the influences of sea-level and buoyancy-loss changes on a hydraulic control model<sup>30</sup> for the Strait of Gibraltar<sup>18,25,31</sup>, which has been independently validated (within uncertainties) by other analytical and numerical solutions<sup>32–35</sup>. The strait model is connected to a basin-representation box model, which includes summer and winter mixed-layer separation, and oxygen isotope fractionation calculations<sup>18,19</sup>. The basin model is identical to that detailed in ref. 19, except that we here omit the so-called ‘monsoon box’, which is relevant only to sapropels (which are excluded here). Following previous Mediterranean habitat identifications<sup>19</sup>, two

$\delta^{18}\text{O}_p$ -to-RSL ‘converters’ are presented: one for (upper) Mediterranean Intermediate Water dweller *Neogloboquadrina pachyderma* (dextral); and one for summer mixed-layer dweller *Globigerinoides ruber* (white). These ‘converters’ (notably that for *G. ruber*) are here used to determine changes in RSL<sub>Gib</sub> from non-sapropelic eastern Mediterranean  $\delta^{18}\text{O}_p$  data (Extended data Fig. 2). Propagated uncertainties in individual RSL<sub>Gib</sub> estimates are up to  $\sim 20$  m ( $1\sigma$ ; see Fig. 1 and Methods), but uncertainty in the mean signal is smaller owing to autocorrelation in the record. A probabilistic assessment that combines RSL<sub>Gib</sub> uncertainties with chronological uncertainties yields a ‘probability maximum’ record with a 95% probability interval of  $\pm 6.3$  m (see Fig. 2 and Methods).

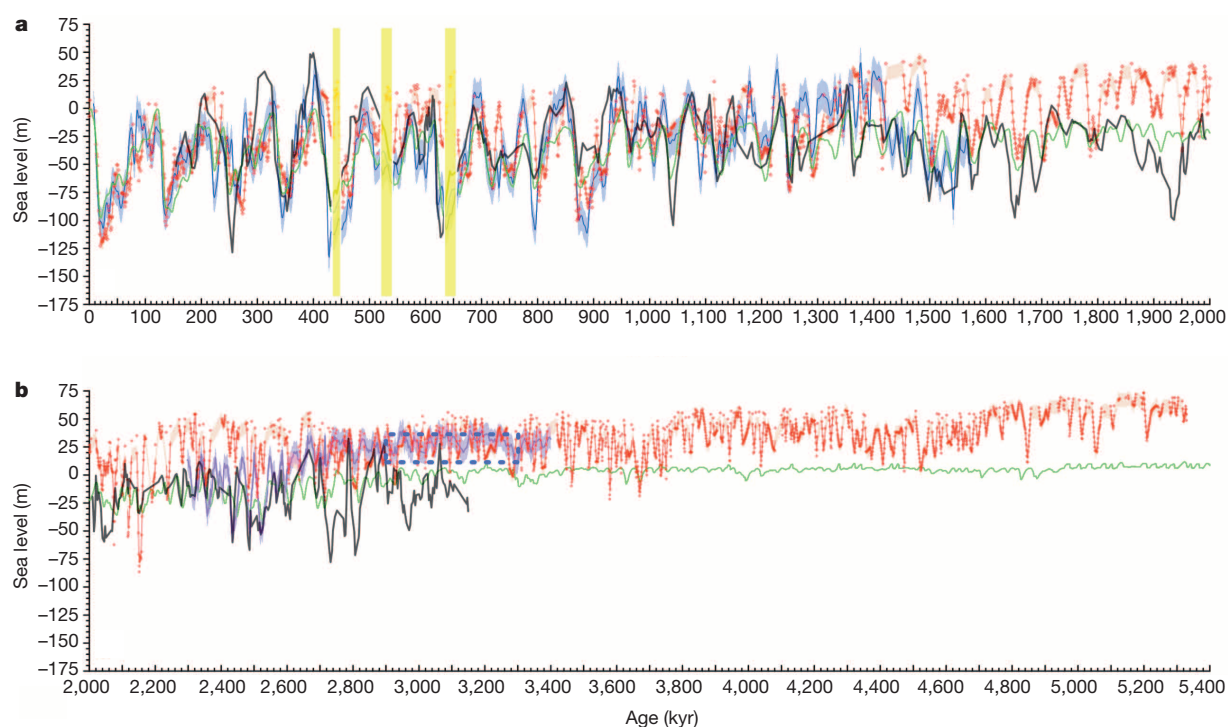
The RSL<sub>Gib</sub> method relies on two underlying assumptions. The first is that the Strait of Gibraltar in the past has exerted hydraulic control on water exchange in a similar manner to today<sup>30</sup>. Large-scale tectonic movement would be detected as a breakdown, or major drift, in the sea-level solutions. Given that the Strait of Gibraltar probably formed during a terminal Miocene event that may have been followed by intense crustal adjustments and erosion<sup>36,37</sup>, it is best to consider the earliest portion of RSL<sub>Gib</sub> with caution. In the interval younger than 3.3 Myr, confidence increases owing to comparison with other sea-level reconstructions (see below), and because of indications that exchange flow through the Strait of Gibraltar adopted a modern-type configuration from about 3.8 Myr ago<sup>38</sup> (Methods). The second assumption concerns our  $\delta^{18}\text{O}$ -to-RSL ‘converter’. It relies on Late Pleistocene parameter relationships with generous uncertainty ranges (Methods), and assumes that past relationships remained within these uncertainty ranges. Again, we expect deviations from this assumption to cause a breakdown, or major drift, in the solutions. Validation between RSL<sub>Gib</sub> and independent methods suggests that this assumption is valid, especially in the past 1.5 Myr and probably throughout (at least) the past 3.3 Myr (Figs 1, 2). Nonetheless,



**Figure 1 | RSL<sub>Gib</sub> compared with RSL<sub>BeM</sub>.** RSL<sub>BeM</sub> has been previously validated against a wide range of independent sea-level benchmarks<sup>11,10,11</sup>. **a**, RSL<sub>Gib</sub> for eastern Mediterranean sediment core LC21 (red) with 1σ error bars (orange), and RSL<sub>BeM</sub> (black) with 2σ error bars<sup>11,10,40</sup> as well as the probabilistically assessed 95% probability envelope (shading)<sup>11</sup>. Individual Mediterranean data comply with the 2σ envelope for the Red Sea data. **b**, RSL<sub>Gib</sub>

for an eastern Mediterranean stack<sup>22</sup> (red) with 1σ error bars (orange), and RSL<sub>BeM</sub> (black) with 2σ error bars<sup>11,10,40</sup>. Gaps in the RSL<sub>Gib</sub> records result from removal of sapropel(-like) events, but some residual influences of freshwater influxes on Mediterranean  $\delta^{18}\text{O}$  (ref. 19) may remain immediately adjacent to these intervals. An apparent ‘undetected’ sapropel-like event (yellow bar) is also indicated (Methods). Note that **a** and **b** cover different age ranges.





**Figure 2 | RSL<sub>Gib</sub> for an eastern Mediterranean  $\delta^{18}\text{O}_p$  stack, compared with other sea-level estimates. a, Interval from 0 to 2 Myr ago. b, Interval from 2 to 5.4 Myr ago. RSL<sub>Gib</sub> (based on the record of ref. 22) is presented using all non-sapropelic data points (red dots), along with the median (red line) from probabilistic analysis with its 95% probability interval (light orange shading). In sapropel intervals, marked by absence of (red) data points, linear interpolation of the 95% probability interval is shown for aesthetic reasons only (Methods). Other sea-level estimates are from Mg/Ca-based  $T_{ds}$ -corrected deep-sea  $\delta^{18}\text{O}_b$  for the southwest Pacific<sup>14</sup> (blue; see Methods for its**

probabilistic presentation here) and for the North Atlantic<sup>15</sup> (black; 3-point moving average); a model-based deconvolution of deep-sea  $\delta^{18}\text{O}_b$  (ref. 12) (green); and conversion of deep-sea  $\delta^{18}\text{O}_b$  with support from New Zealand sequence stratigraphic data<sup>23</sup> (purple). The last was vertically positioned to agree with the 12–32 m ESL estimate for the period 2.9–3.3 Myr ago (PLIOMAX data presented at PALSEA2 meeting; M. E. Raymo, personal communication; dark blue dashed box). Three apparent ‘undetected’ sapropel-like intervals in RSL<sub>Gib</sub> are indicated (yellow bars; Methods).

given the potential caveats to our method with respect to long-term tectonic and climate-regime changes, we emphasize the need for new continuous and highly resolved records from independent methods to strengthen mutual validations before 1.5 Myr ago, and especially before 3.3 Myr ago (Methods).

We also make an initial assessment of land movement due to glacio-hydro-isostatic adjustment at the Camarinal sill, which is the shallowest and hydraulically limiting passage on the Atlantic side of the Gibraltar narrows<sup>30</sup> (Methods). We find that the glacial–interglacial amplitude of RSL<sub>Gib</sub> underestimates global mean (eustatic) sea level (hereafter ESL), and that the offset scales proportionally with the variation in land ice (where  $\text{ESL} = 1.23 \text{ RSL}_{\text{Gib}}$ ; see Methods), so that ESL amplitude variations will be larger than those of RSL<sub>Gib</sub>. This scaling suggests Pliocene isostatic adjustments in the region that are compatible with previous estimates<sup>39</sup>. Regardless, in our comparisons with estimates from other methods (Figs 1 and 2), we plot RSL<sub>Gib</sub> instead of ESL, because issues about extension of our isostatic assessment back in time, and to periods with sea level considerably above the present level, remain to be constrained. We note that the international ‘Pliocene Maximum Sea Level’ (PLIOMAX) project currently estimates ESL at 12–32 m for the period 3.3–2.9 Myr ago (M. E. Raymo, personal communication), similar to the Pliocene range of 9–31 m used in ref. 4. These values agree well with RSL<sub>Gib</sub> fluctuations during that period (Fig. 2). PLIOMAX estimates do not specify the nature and magnitude of temporal variability, so our RSL<sub>Gib</sub> record provides the first quantitative view of the secular evolution of sea level during the Pliocene that is independent of deep-sea  $\delta^{18}\text{O}_b$ .

### RSL<sub>Gib</sub> validation

Where overlap exists, comparison of RSL<sub>Gib</sub> with independent sea-level reconstructions reveals good agreement (Figs 1 and 2). For younger intervals (<0.5 Myr ago), we now assess RSL<sub>Gib</sub> values derived from

different  $\delta^{18}\text{O}_p$  data sets on their respective timescales. RSL<sub>Gib</sub> from  $\delta^{18}\text{O}_p$  in eastern Mediterranean sediment core LC21—which has a closely related chronology to RSL<sub>BeM</sub><sup>11</sup>—agrees well with RSL<sub>BeM</sub> over the last glacial cycle through intervals of both deglaciation (145–125 kyr ago) and glacial inception (120–65 kyr ago), and has the additional benefit of resolving changes between 26 and 14 kyr ago, where RSL<sub>BeM</sub> is poorly resolved due to aplanktonic conditions in the Red Sea (Fig. 1a). Using an eastern Mediterranean  $\delta^{18}\text{O}_p$  stack<sup>22</sup>, and allowing for different age models, RSL<sub>Gib</sub> also compares well with RSL<sub>BeM</sub> over the past 0.5 Myr (refs 1, 40; Fig. 1b). Agreement between the two RSL records from different ocean margins reflects the fact that both areas have generally comparable glacio-hydro-isostatic responses (Methods). The principal disagreements lie close to sapropel(-like) intervals of freshwater dilution in the Mediterranean, which tend to bias RSL<sub>Gib</sub> towards higher values. This suggests that residual effects of these events may occasionally remain, owing to imperfect detection/removal of sapropel(-like) intervals (Methods). Future work can use same-sample multi-proxy approaches to improve this situation, and reinstate the ‘monsoon box’ in the model<sup>19</sup> to try and resolve RSL<sub>Gib</sub> through sapropel intervals, but these are multi-year efforts beyond the scope of the present study (Methods).

Next we compare the full RSL<sub>Gib</sub> record with available sea-level reconstructions for older (>0.5 Myr ago) intervals (Fig. 2). We observe strong agreement between RSL<sub>Gib</sub> and sea-level estimates from Mg/Ca  $T_{ds}$ -corrected deep-sea  $\delta^{18}\text{O}_b$  over the past 1.5 Myr (ref. 14; Fig. 2a), which independently supports the intensification of glacials across the Mid-Pleistocene transition that was first inferred from the southwest Pacific  $T_{ds}$ -corrected deep-sea  $\delta^{18}\text{O}_b$  record<sup>14</sup>. However, before ~1.5 Myr ago, highstand values of RSL<sub>Gib</sub> seem higher than those in other studies<sup>14,15</sup>. Apart from RSL<sub>Gib</sub>, the only records with continuity across the past 3 Myr are an Atlantic  $T_{ds}$ -corrected deep-sea  $\delta^{18}\text{O}_b$  record<sup>15</sup>, and the model-based estimates of ref. 12. Relative to RSL<sub>Gib</sub>, both suggest a lower mean

sea level between  $\sim 1.5$  and  $3.2$  Myr ago (Fig. 2). If we assume that this difference arises from bias in  $\text{RSL}_{\text{Gib}}$ , then it might reflect a more open strait before  $\sim 1.5$  Myr ago, possibly due to an uplift event in the Strait of Gibraltar at  $\sim 1.5$  Myr ago. However, such an event would also affect the sea-level sensitivity (amplitude response) of Mediterranean  $\delta^{18}\text{O}$ . This is difficult to reconcile with the observation that  $\text{RSL}_{\text{Gib}}$  amplitude variations agree well with those in a record based on scaling of a New Zealand sequence stratigraphic record between  $3.4$  and  $2.3$  Myr ago<sup>2,3</sup>, and also with those in the Atlantic record (allowing for resolution and chronological differences)<sup>15</sup> (Fig. 2b). Hence, we suggest that any change in  $\text{RSL}_{\text{Gib}}$  at  $\sim 1.5$  Myr ago is more likely to reflect a 'baseline shift' in Mediterranean climate conditions from a warm/moist state to a warm/arid state, rather than a tectonic step at the Strait of Gibraltar. Alternatively, we might assume that the bias is not due to  $\text{RSL}_{\text{Gib}}$ , given that there is good agreement between  $\text{RSL}_{\text{Gib}}$  and PLIOMAX estimates for the  $2.9$ – $3.3$  Myr interval (Fig. 2b). That would suggest that problems may instead lie within the Atlantic  $T_{\text{ds}}$ -corrected deep-sea  $\delta^{18}\text{O}_{\text{b}}$  record<sup>15</sup> (as also suggested before; see, for example, ref. 16), and within the model-based estimates of ref. 12. Further independent validation is needed before this can be settled, but—regardless—major signal-amplitude similarity in all independent observational methods over the past  $3.3$  Myr challenges the substantially different sea-level inferences from model-based deconvolution of deep-sea  $\delta^{18}\text{O}_{\text{b}}$  (ref. 12; Fig. 2a, b).

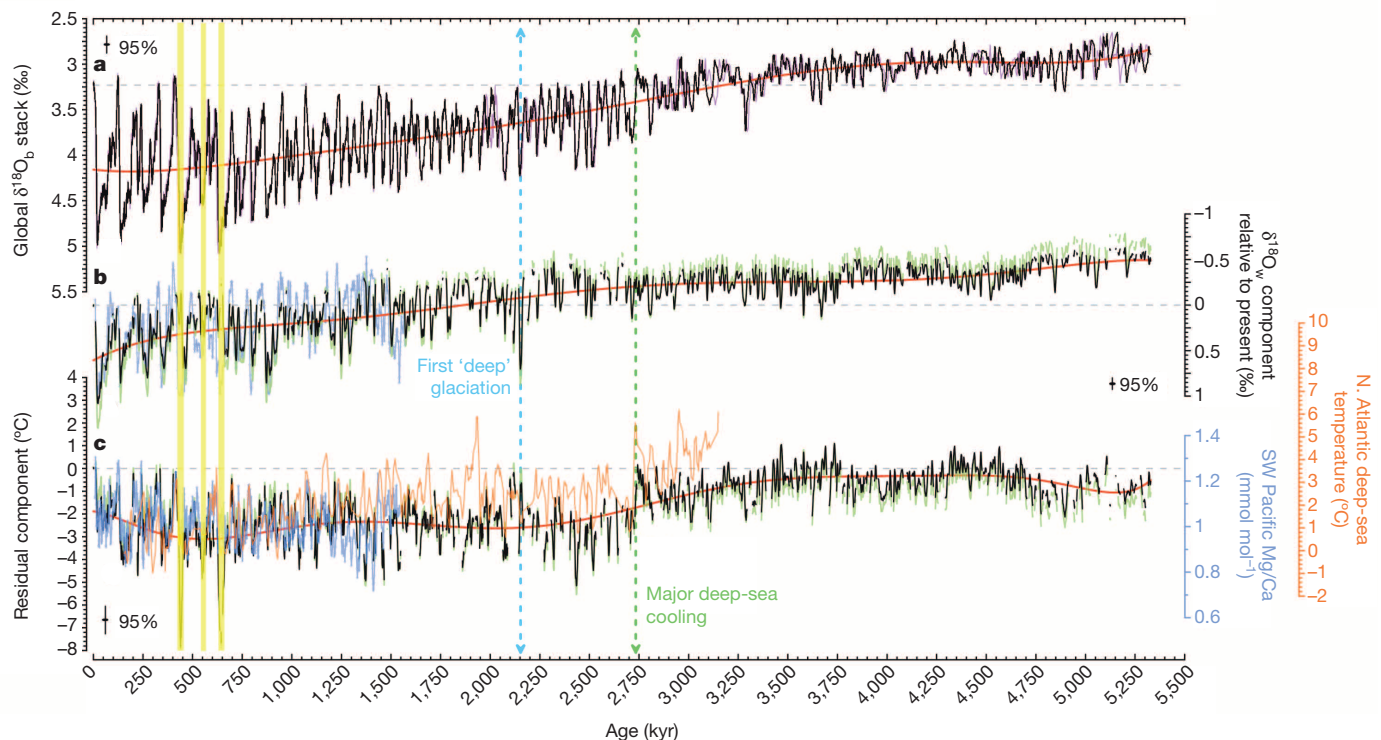
### Deep-water temperature variability

Next, we use  $\text{RSL}_{\text{Gib}}$  to derive information about global  $T_{\text{ds}}$  changes. Owing to uncertainty in the long-term  $\text{RSL}_{\text{Gib}}$ :ESL scaling, we consider two scenarios, one of which relies on direct use of  $\text{RSL}_{\text{Gib}}$  and the other

on an ESL estimate of  $1.23 \times \text{RSL}_{\text{Gib}}$  (Figs 3, 4). We translate these into estimates of seawater  $\delta^{18}\text{O}$  ( $\delta^{18}\text{O}_{\text{w}}$ ) changes using a ratio of  $(0.009 \pm 0.001)\text{‰ m}^{-1}$  (refs 14, 41, 42; Figs 3b and 4b). Subtraction of  $\delta^{18}\text{O}_{\text{w}}$  from a global deep-sea  $\delta^{18}\text{O}_{\text{b}}$  stack<sup>43</sup> (Figs 3a and 4a), following slight adjustment of the chronology of this stack to that of the  $\text{RSL}_{\text{Gib}}$  record (Methods; Extended Data Table 1), yields residuals that approximate global  $T_{\text{ds}}$  changes in a  $0.25\text{‰ °C}^{-1}$  ratio<sup>14</sup>. Thus, we estimate global  $T_{\text{ds}}$  changes over the past  $5.3$  Myr, with propagated ( $2\sigma$ ) uncertainties of about  $\pm 0.6\text{ °C}$  (Figs 3c and 4c). Our estimates agree well with independent Mg/Ca-based  $T_{\text{ds}}$  estimates for a site that is thought to approximate global mean deep-water conditions<sup>14</sup> (Figs 3c and 4c). The observed mutual consistency over glacial–interglacial cycles and longer timescales between this Mg/Ca-based  $T_{\text{ds}}$  record and associated sea-level reconstruction<sup>14</sup>, and our  $\text{RSL}_{\text{Gib}}$  and associated  $T_{\text{ds}}$  reconstruction, suggests that over the past  $1.6$  Myr (1) seawater Mg/Ca ratios did not change significantly and (2) Strait of Gibraltar morphology and the Mediterranean 'baseline climate state' experienced no major changes. In addition, the lower-resolution Atlantic  $T_{\text{ds}}$  record<sup>15</sup> also validates a major  $T_{\text{ds}}$  drop in our reconstruction at  $\sim 2.73$  Myr ago (see below; Fig. 3c).

### Timing and magnitude of glaciations

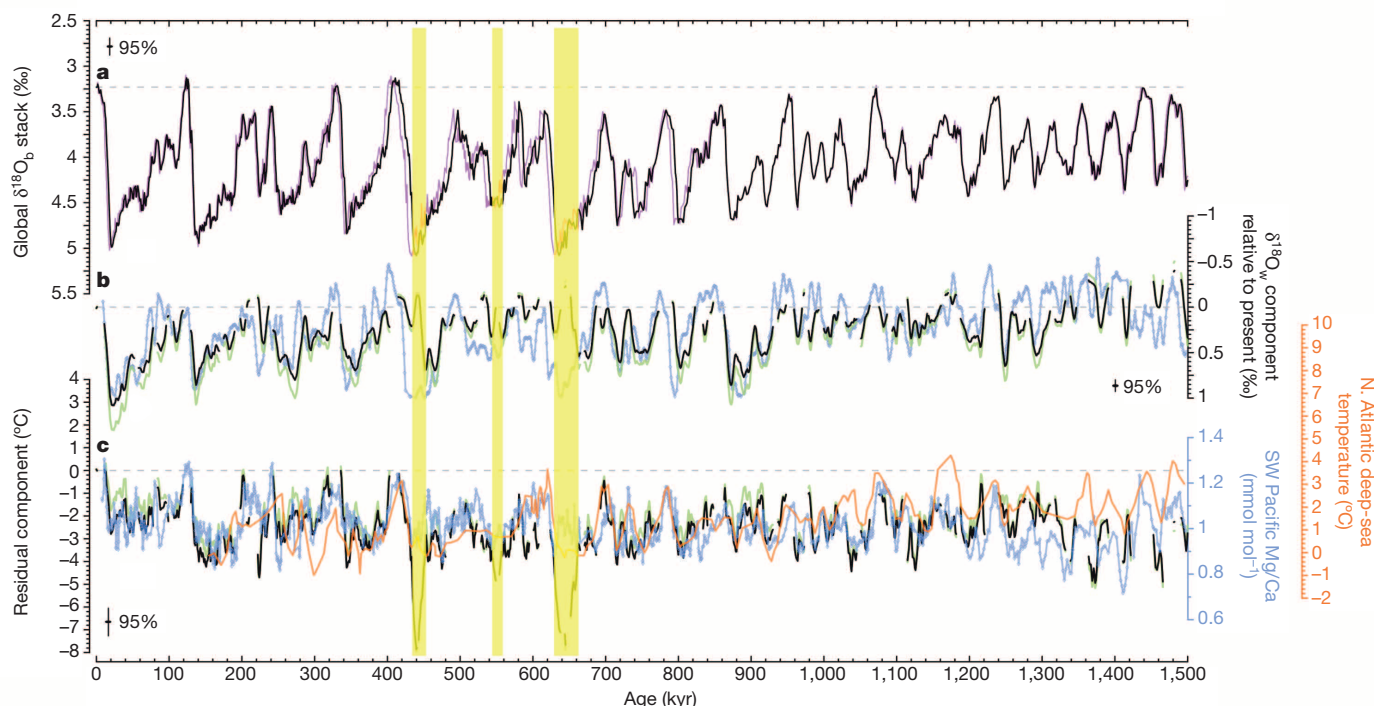
Several key observations can be drawn from our analysis. First, we extend to  $5.3$  Myr ago the conclusion, previously drawn for the past  $1.5$  Myr (ref. 14), that global deep-sea  $\delta^{18}\text{O}$  (see, for example, refs 13, 43) does not adequately capture ice-volume history because its two main components ( $T_{\text{ds}}$  and ice-volume effects) underwent distinctly different temporal developments (Fig. 3). Second, regarding the onset of Quaternary glacial cycles, we find that a distinct deep-sea cooling step at  $2.73$  Myr



**Figure 3 | Deep-sea temperature and  $\delta^{18}\text{O}_{\text{w}}$  components of deep-sea  $\delta^{18}\text{O}_{\text{b}}$ .** **a**, Deep-sea  $\delta^{18}\text{O}_{\text{b}}$  (ref. 43) on original chronology (magenta) and chronology tuned to that of ref. 22 (black). **b**, Component of sea-level-based ocean  $\delta^{18}\text{O}_{\text{w}}$  variations (in black based on  $\text{RSL}_{\text{Gib}}$  and in green for our ESL approximation, both using  $0.009\text{‰ m}^{-1}$ ), compared with  $\delta^{18}\text{O}_{\text{w}}$  for the southwest Pacific<sup>14</sup> (blue; 3-point moving average). All variations are assessed relative to present. **c**, Residual  $\delta^{18}\text{O}$  component that is ascribed to  $T_{\text{ds}}$  changes using  $0.25\text{‰ °C}^{-1}$ . Gaps in **b** and **c** relate to sapropel(-like) intervals. For details see Methods. Also shown in **c** are 3-point moving averages of Mg/Ca-based  $T_{\text{ds}}$  records for the North Atlantic<sup>15</sup> (orange) and the southwest Pacific (blue)<sup>14</sup>. The southwest Pacific record is shown in original Mg/Ca units (as made available), but is

exactly scaled to  $T_{\text{ds}}$  variations on the other axes as described in ref. 14. Error bars: **a**, mean uncertainties reported for deep-sea  $\delta^{18}\text{O}_{\text{b}}$  in the global stack<sup>43</sup>; **b**, propagated uncertainty in the sea-level ( $\text{RSL}_{\text{Gib}}$ )-based  $\delta^{18}\text{O}_{\text{w}}$  change component, based on the 95% probability envelope to the median (Methods, and Fig. 2) and a  $\pm 0.001\text{‰ m}^{-1}$  uncertainty in the conversion to  $\delta^{18}\text{O}_{\text{w}}$ ; and **c**, propagated uncertainties from **a** and **b**. Yellow bars as in Fig. 2. A major deep-sea cooling (green dashed line), and the first 'deep' glaciation (sea-level lowering below  $-70$  m; blue dashed line) are indicated. Red lines are straightforward polynomial fits shown only to highlight general long-term trends (all based directly on  $\text{RSL}_{\text{Gib}}$ ). Age on x axis is based on the chronology of ref. 22. A magnification of the past  $1.5$  Myr is shown in Fig. 4.





**Figure 4 | Expanded version of Fig. 3 for the past 1.5 Myr only.** a–c, As in Fig. 3. Error bars: a, mean uncertainties reported for deep-sea  $\delta^{18}\text{O}_b$  in the global stack<sup>43</sup>; b, propagated uncertainty in the sea-level (RSL<sub>Gib</sub>)-based  $\delta^{18}\text{O}_w$

change component, based on the 95% probability envelope to the median (Methods, and Fig. 2) and a  $\pm 0.001\text{‰ m}^{-1}$  uncertainty in the conversion to  $\delta^{18}\text{O}_w$ ; and c, propagated uncertainties from a and b.

ago substantially pre-dated the first major glaciation in our record, by 0.58 Myr (Fig. 3). Note that bias due to sapropel intervals is towards high RSL<sub>Gib</sub> values, and that comparisons between well-defined RSL<sub>Gib</sub> lowstand values are robust relative to this bias. This strengthens confidence in our identification of the first major lowstand, especially because data for the event at 2.15 Myr ago appear to be  $\sim 60\text{ m}$  ( $3\sigma$ ) lower than for any preceding lowstand (Fig. 2).

Current concepts for the onset of Northern Hemisphere glaciation rely strongly on deep-sea  $\delta^{18}\text{O}_b$  data, and suggest a shift to stronger glacials at  $\sim 2.7\text{--}2.5\text{ Myr ago}$  (refs 43, 44; Fig. 3a). Our new data challenge this perspective, because the change at 2.73 Myr ago appears to relate to cooling, whereas the first ‘deep’ (sea level below  $\sim 70\text{ m}$ ) glacial occurred considerably later, at 2.15 Myr ago. Pronounced cooling at  $\sim 2.73\text{ Myr ago}$  is supported not only by the independent Atlantic deep-sea Mg/Ca record<sup>15</sup> (Fig. 3c), but also by an alkenone-based North Atlantic surface temperature record<sup>6</sup>, and the culmination of a long-term equatorial Pacific cooling trend<sup>45</sup>. It is consistent with ample evidence for widespread ocean and climate change at  $\sim 2.7\text{ Myr ago}$  (Extended Data Table 2), including glaciation on Greenland and Scandinavia (see synthesis in ref. 8). Apparent temporal association of this cooling with a decline in atmospheric  $\text{CO}_2$  levels (see, for example, ref. 46) suggests a causal link. For instance, stratification in both the North Pacific and the Southern Ocean intensified at  $\sim 2.7\text{ Myr ago}$  in association with deep-sea cooling, leading to increased ocean carbon storage<sup>5</sup>. These are key regions of deep and intermediate water formation (particularly the Southern Ocean), and changes in their overturning circulation may, therefore, strongly influence widespread oceanic carbon storage and, hence, atmospheric  $\text{CO}_2$  levels (see, for example, ref. 47). Our inferred first ‘deep’ glacial at 2.15 Myr ago also falls within a window of major climatic and oceanic changes (Extended Data Table 2), including a major cooling in tropical sea surface temperatures<sup>48</sup>, but its nature requires further validation (for example, through extension of the record of ref. 14).

Terrestrial indications that a major North American ice sheet developed to low latitudes ( $39^\circ\text{ N}$ ) date to  $\sim 2.4\text{ Myr ago}$ , while the earliest

record of significant North American-sourced ice-rafted debris suggests that ice sheets extended to marine margins at 2.64 Myr ago (refs 7, 8). RSL<sub>Gib</sub> has amplitudes of  $50\text{--}70\text{ m}$  at that time (Fig. 2b), which imply that early ice sheets had relatively low profiles relative to their large inferred areas. Such ‘low-slung’ ice sheets may have existed because basal friction was lower during early glacial cycles than during more recent ones<sup>49,50</sup>.

Finally, we infer that similar amplitudes among three of the last four glacial maxima in the global deep-sea  $\delta^{18}\text{O}_b$  stack<sup>43</sup> (Figs 3a and 4a) may obscure increasing ice-volume contributions (Figs 3b and 4b) that are compensated by decreasing deep-sea temperature contributions (Figs 3c and 4c). Two independent methods (this Article and ref. 14) suggest that the Last Glacial Maximum was one of the most intense glaciations in terms of ice volume, but that its deep-sea temperatures may have been relatively ‘mild’ by glacial standards (Figs 3 and 4). This apparent difference in  $T_{ds}$  between glacials is smaller in the southwest Pacific record<sup>14</sup> than in our reconstruction (Fig. 4); possibly, it was most notable in the Atlantic Ocean, which dominates the  $\delta^{18}\text{O}_b$  stack<sup>43</sup> that we used to estimate  $T_{ds}$ . The pattern also is less evident in RSL<sub>BeM</sub>, but RSL<sub>BeM</sub> is known to be deficient through the LGM<sup>10,17</sup> (Fig. 1). At this stage, therefore, the inferred pattern is suggestive only; it requires validation from both improved RSL<sub>Gib</sub> reconstructions based on continuous multi-proxy core records, and additional deep-sea benthic Mg/Ca records. If validated, then it may reflect the importance of atmospheric moisture supply (and reduced atmospheric moisture capacity with decreasing temperature) in determining total ice accumulation.

## METHODS SUMMARY

In Methods, we explain (1) our Mediterranean relative sea level at Gibraltar (RSL<sub>Gib</sub>) calculations, including elimination of so-called sapropelic intervals with freshwater dilution, a discussion of long-term tectonic effects, an assessment of glacio-hydro-isostatic influences and an outline of scope for future refinements; (2) our probabilistic assessment of the sea-level record of ref. 14; and (3) our use of RSL<sub>Gib</sub> with the deep-sea  $\delta^{18}\text{O}$  stack of ref. 43 to determine changes in deep-sea temperature ( $T_{ds}$ ). A full copy of our PTC MathCad 13 worksheet, which was used

to calculate the RSL<sub>Gib</sub> relationship with eastern Mediterranean  $\delta^{18}\text{O}$ , as measured on the planktonic foraminiferal species *Globigerinoides ruber* (white) and *Neogloboquadrina pachyderma* (dextral), is available in Supplementary Information.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 13 November 2013; accepted 4 March 2014.**

**Published online 16 April 2014.**

- Rohling, E. J. *et al.* Antarctic temperature and global sea level closely coupled over the past five glacial cycles. *Nature Geosci.* **2**, 500–504 (2009).
- Naish, T. *et al.* Obliquity-paced Pliocene West Antarctic ice sheet oscillations. *Nature* **458**, 322–328 (2009).
- Miller, K. G. *et al.* High tide of the warm Pliocene: implications of global sea level for Antarctic deglaciation. *Geology* **40**, 407–410 (2012).
- Foster, G. L. & Rohling, E. J. The relationship between sea level and climate forcing by  $\text{CO}_2$  on geological timescales. *Proc. Natl Acad. Sci. USA* **110**, 1209–1214 (2013).
- Sigman, D. M., Jaccard, S. & Haug, G. H. Polar ocean stratification in a cold climate. *Nature* **428**, 59–63 (2004).
- Lawrence, K. T., Herbert, T. D., Brown, C. M., Raymo, M. E. & Haywood, A. M. High amplitude variations in North Atlantic sea surface temperature during the early Pliocene warm period. *Paleoceanography* **24**, PA2218 (2009).
- Balco, G. & Rovey, C. W. II Absolute chronology for major Pleistocene advances of the Laurentide ice sheet. *Geology* **38**, 795–798 (2010).
- Bailey, I. *et al.* An alternative suggestion for the Pliocene onset of major northern hemisphere glaciation based on the geochemical provenance of North Atlantic Ocean ice-rafted debris. *Quat. Sci. Rev.* **75**, 181–194 (2013).
- Waelbroeck, C. *et al.* Sea-level and deep water temperature changes derived from benthic foraminifera isotopic records. *Quat. Sci. Rev.* **21**, 295–305 (2002).
- Siddall, M. *et al.* Sea-level fluctuations during the last glacial cycle. *Nature* **423**, 853–858 (2003).
- Grant, K. M. *et al.* Rapid coupling between ice volume and polar temperature over the past 150 kyr. *Nature* **491**, 744–747 (2012).
- de Boer, B., Van de Wal, R. S. W., Bintanja, R., Lourens, L. J. & Tuenner, E. Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic  $\delta^{18}\text{O}$  records. *Ann. Glaciol.* **51**, 23–33 (2010).
- Miller, K. G. *et al.* The Phanerozoic record of global sea-level change. *Science* **310**, 1293–1298 (2005).
- Elderfield, H. *et al.* Evolution of ocean temperature and ice volume through the Mid-Pleistocene Climate Transition. *Science* **337**, 704–709 (2012).
- Sosdian, S. & Rosenthal, Y. Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions. *Science* **325**, 306–310 (2009).
- Yu, J. & Broecker, W. S. Comment on “Deep-sea temperature and ice volume changes across the Pliocene-Pleistocene climate transitions”. *Science* **328**, 1480C (2010).
- Siddall, M. *et al.* Understanding the Red Sea response to sea level. *Earth Planet. Sci. Lett.* **225**, 421–434 (2004).
- Rohling, E. J. Environmental controls on salinity and  $\delta^{18}\text{O}$  in the Mediterranean. *Paleoceanography* **14**, 706–715 (1999).
- Rohling, E. J. *et al.* Reconstructing past planktic foraminiferal habitats using stable isotope data: a case history for Mediterranean sapropel S5. *Mar. Micropaleontol.* **50**, 89–123 (2004).
- Lourens, L. J. *et al.* Evaluation of the Plio-Pleistocene astronomical timescale. *Paleoceanography* **11**, 391–413 (1996).
- Emeis, K.-C., Sakamoto, T., Wehausen, R. & Brumsack, H.-J. The sapropel record of the eastern Mediterranean Sea — results of Ocean Drilling Program Leg 160. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **158**, 371–395 (2000).
- Wang, P., Tian, J. & Lourens, L. J. Obscuring of long eccentricity cyclicity in Pleistocene oceanic carbon isotope records. *Earth Planet. Sci. Lett.* **290**, 319–330 (2010).
- Lourens, L. J., Wehausen, R. & Brumsack, H.-J. Geological constraints on tidal dissipation and dynamical ellipticity of the Earth over the past three million years. *Nature* **409**, 1029–1033 (2001).
- Rohling, E. J. Quantitative assessment of glacial fluctuations in the level of Lake Lisan, Dead Sea rift. *Quat. Sci. Rev.* **70**, 63–72 (2013).
- Rohling, E. J. Review and new aspects concerning the formation of Mediterranean sapropels. *Mar. Geol.* **122**, 1–28 (1994).
- Jorissen, F. J. Benthic foraminiferal successions across Late Quaternary Mediterranean sapropels. *Mar. Geol.* **153**, 91–101 (1999).
- Thomson, J., Mercione, D., De Lange, G. J. & Van Santvoort, P. J. M. Review of recent advances in the interpretation of eastern Mediterranean sapropel S1 from geochemical evidence. *Mar. Geol.* **153**, 77–89 (1999).
- Mercione, D. *et al.* High-resolution geochemical and micropalaeontological probing of the most recent eastern Mediterranean sapropel. *Mar. Geol.* **177**, 25–44 (2001).
- De Lange, G. J. *et al.* Synchronous basin-wide formation and redox-controlled preservation of a Mediterranean sapropel. *Nature Geosci.* **1**, 606–610 (2008).
- Bryden, H. L. & Kinder, T. H. Steady two-layer exchange through the Strait of Gibraltar. *Deep Sea Res.* **38**, S445–S463 (1991).
- Rohling, E. J. & Bryden, H. L. Estimating past changes in the eastern Mediterranean freshwater budget, using reconstructions of sea level and hydrography. *Proc. Kon. Ned. Akad. B* **97**, 201–217 (1994).
- Myers, P., Haines, K. & Rohling, E. J. Modelling the paleo-circulation of the Mediterranean: the last glacial maximum and the Holocene with emphasis on the formation of sapropel S1. *Paleoceanography* **13**, 586–606 (1998).
- Matthies, S. & Haines, K. A hydraulic box model study of the Mediterranean response to postglacial sea-level rise. *Paleoceanography* **18**, 1084 (2003).
- Mikolajewicz, U. Modelling Mediterranean ocean climate of the Last Glacial Maximum. *Clim. Past* **7**, 161–180 (2011).
- Rogerson, M., Bigg, G. R., Rohling, E. J. & Ramirez, J. Vertical density gradient in the eastern North Atlantic during the last 30,000 years. *Clim. Dyn.* **39**, 589–598 (2012).
- Loget, N. & Van der Driessche, J. On the origin of the Strait of Gibraltar. *Sedim. Geol.* **188–189**, 341–356 (2006).
- García-Castellanos, D. & Villaseñor, A. Messinian salinity crisis regulated by competing tectonics and erosion at the Gibraltar arc. *Nature* **480**, 359–363 (2011).
- Hernández-Molina, F. J. *et al.* Contourite processes associated with the Mediterranean outflow water after its exit from the Strait of Gibraltar: global and conceptual implications. *Geology* **42**, 227–230 (2014).
- Raymo, M. E., Mitrovica, J. X., O’Leary, M. J., DeConto, R. M. & Hearty, P. J. Departures from eustasy in Pliocene sea-level records. *Nature Geosci.* **4**, 328–332 (2011).
- Rohling, E. J. *et al.* Comparison between Holocene and Marine Isotope Stage-11 sea-level histories. *Earth Planet. Sci. Lett.* **291**, 97–105 (2010).
- Schrag, D. P., Hampt, G. & Murray, D. W. Pore fluid constraints on the temperature and oxygen isotopic composition of the glacial ocean. *Science* **272**, 1930–1932 (1996).
- Adkins, J. F., McIntyre, K. & Schrag, D. P. The salinity, temperature, and  $\delta^{18}\text{O}$  of the glacial deep ocean. *Science* **298**, 1769–1773 (2002).
- Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, PA1003 <http://dx.doi.org/10.1029/2004PA001071> (2005).
- Shackleton, N. J. & Opdyke, N. D. Oxygen isotope and palaeomagnetic evidence for early Northern Hemisphere glaciation. *Nature* **270**, 216–219 (1977).
- Medina-Elizalde, M., Lea, D. W. & Fiant, M. S. Implications of seawater Mg/Ca variability for Plio-Pleistocene tropical climate reconstruction. *Earth Planet. Sci. Lett.* **269**, 585–595 (2008).
- Seki, O. *et al.* Alkenone and boron-based Pliocene  $p\text{CO}_2$  records. *Earth Planet. Sci. Lett.* **292**, 201–211 (2010).
- Sigman, D. M., Hain, M. P. & Haug, G. H. The polar ocean and glacial cycles in atmospheric  $\text{CO}_2$  concentration. *Nature* **466**, 47–55 (2010).
- Herbert, T. D., Cleaveland Peterson, L., Lawrence, K. T. & Liu, Z. Tropical ocean temperatures over the past 3.5 million years. *Science* **328**, 1530–1534 (2010).
- Clark, P. U. & Pollard, D. Origin of the middle Pleistocene transition by ice sheet erosion of regolith. *Paleoceanography* **13**, 1–9 (1998).
- Bailey, I. *et al.* A low threshold for North Atlantic ice rafting from “low-slung slippery” late Pliocene ice sheets. *Paleoceanography* **25**, PA1212 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Raymo for discussion of Pliocene sea-level estimates (notably PLIOMAX) at the PALSEA2 workshop in Rome, October 2013, and all colleagues who made their data available—for example, via the PANGAEA and NOAA-NCDC Palaeoclimate data centres, or directly. This study was supported by 2012 Australian Laureate Fellowship FL120100050 (E.J.R.) and UK Natural Environment Research Council (NERC) consortium project iGlass (E.J.R., M.T., F.W., A.P.R.). F.W. acknowledges an Australian Bicentennial Scholarship Award from the Menzies Centre for Australian Studies, King’s College London.

**Author Contributions** E.J.R. led the study, and performed the calculations. F.W. contributed the assessment of isostatic effects under the guidance of M.T. All authors contributed specialist insights to the discussions and helped with composing and refining the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.J.R. ([eelco.rohling@anu.edu.au](mailto:eelco.rohling@anu.edu.au)).



# Juno is the egg Izumo receptor and is essential for mammalian fertilization

Enrica Bianchi<sup>1</sup>, Brendan Doe<sup>2</sup>, David Goulding<sup>3</sup> & Gavin J. Wright<sup>1</sup>

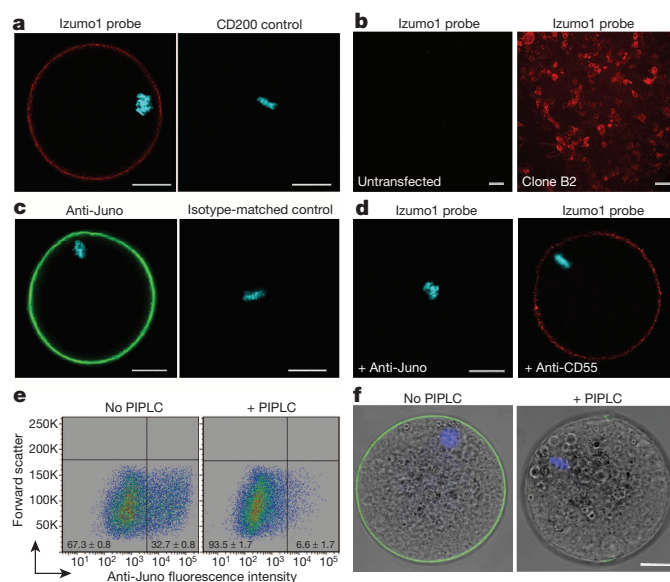
**Fertilization occurs when sperm and egg recognize each other and fuse to form a new, genetically distinct organism. The molecular basis of sperm–egg recognition is unknown, but is likely to require interactions between receptor proteins displayed on their surface. Izumo1 is an essential sperm cell–surface protein, but its receptor on the egg has not been described. Here we identify folate receptor 4 (Folr4) as the receptor for Izumo1 on the mouse egg, and propose to rename it Juno. We show that the Izumo1–Juno interaction is conserved within several mammalian species, including humans. Female mice lacking *Juno* are infertile and *Juno*-deficient eggs do not fuse with normal sperm. Rapid shedding of Juno from the oolemma after fertilization suggests a mechanism for the membrane block to polyspermy, ensuring eggs normally fuse with just a single sperm. Our discovery of an essential receptor pair at the nexus of conception provides opportunities for the rational development of new fertility treatments and contraceptives.**

Fertilization is the culminating event in sexual reproduction and requires the fusion of haploid sperm and egg to create a new, genetically distinct, diploid organism. Sperm acquire the ability to fertilize the egg within the female reproductive tract by exposing previously concealed receptor proteins onto their surface following the acrosome reaction<sup>1</sup>. Once fertilized, both the oolemma and zona pellucida are biochemically altered, making the egg unreceptive to additional sperm and thereby reducing the chances of creating nonviable polyploid embryos<sup>2</sup>. Several receptor proteins have been implicated in the recognition and/or fusion process<sup>3</sup>, but just two significantly affect fertility *in vivo*: Izumo1 on sperm<sup>4</sup>, and CD9 on eggs<sup>5–7</sup>. Izumo1 (named after a Japanese marriage shrine) is redistributed to the surface of capacitated sperm<sup>8</sup>, and *Izumo1*-deficient male—but not female—mice are infertile because sperm lacking Izumo1 cannot fuse with eggs<sup>4</sup>. Recombinant Izumo1 binds both wild-type and CD9-deficient eggs, suggesting that Izumo1 interacts with an egg receptor other than CD9<sup>9</sup>. Glycophosphatidylinositol (GPI)-anchored receptors on the egg are essential for fertilization because removing them either enzymatically<sup>10</sup> or genetically<sup>11</sup> renders eggs infertile. Despite these advances, the molecular basis of gamete recognition in mammals is unknown; in part, this is due to the scarcity of eggs, the challenges in solubilising membrane-embedded proteins, and the often transient nature of their extracellular interactions<sup>12</sup>. To address these challenges, we have developed techniques to identify low-affinity extracellular interactions and apply them here to investigate sperm–egg recognition in mammals. We now describe the Izumo1 egg receptor as folate receptor 4 (Folr4), a GPI-anchored protein expressed on the egg surface that is essential for female fertility. Because of its role in fertilization and inability to bind folate, we have renamed this protein ‘Juno’ after the Roman goddess of fertility and marriage.

## Juno is the Izumo1 receptor on oocytes

To identify an egg receptor for Izumo1, we expressed the entire ectodomain of mouse Izumo1 in mammalian cells for use as a binding probe. With the expectation that interactions with its extracellular egg receptor would be weak, we oligomerized the Izumo1 ectodomain to increase binding avidity by using a peptide sequence from a cartilage protein that forms pentamers<sup>12</sup> (Extended Data Fig. 1). The avid Izumo1 probe,

but not a control protein, bound the oolemma of mouse oocytes (Fig. 1a). To determine the molecular identity of the egg binding partner for Izumo1, we used a mouse oocyte complementary DNA library in an iterative expression cloning approach (Extended Data Fig. 2). In each round of selection, an increasing fraction of transfected cells bound Izumo1 until three independently selected plasmid clones conferred Izumo1 binding (Fig. 1b). All three clones contained the same cDNA, encoding the folate



**Figure 1 | Juno is the GPI-anchored oocyte surface receptor for Izumo1.**

**a**, An avid recombinant Izumo1 protein, but not a control, binds the oolemma. **b**, Izumo1 binds HEK293 cells transfected with a *Juno* cDNA (clone B2), but not untransfected controls. **c**, Juno is highly expressed on the oolemma of unfertilized eggs. **d**, Preincubation of eggs with an anti-Juno, but not an anti-CD55 control antibody, blocked binding of recombinant Izumo1 protein. **e**, **f**, Cell-surface Juno was lost after treatment with PIPLC on *Juno*-transfected HEK293 cells (**e**) or eggs (**f**). Images are single optical sections of unfertilized mouse metaphase II eggs; scale bars, 20  $\mu$ m in **a**, **c**, **d**, **f**, and 10  $\mu$ m in **b**.

<sup>1</sup>Cell Surface Signalling Laboratory, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. <sup>2</sup>Mouse Production Team, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. <sup>3</sup>Electron and Advanced Light Microscopy Suite, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK.

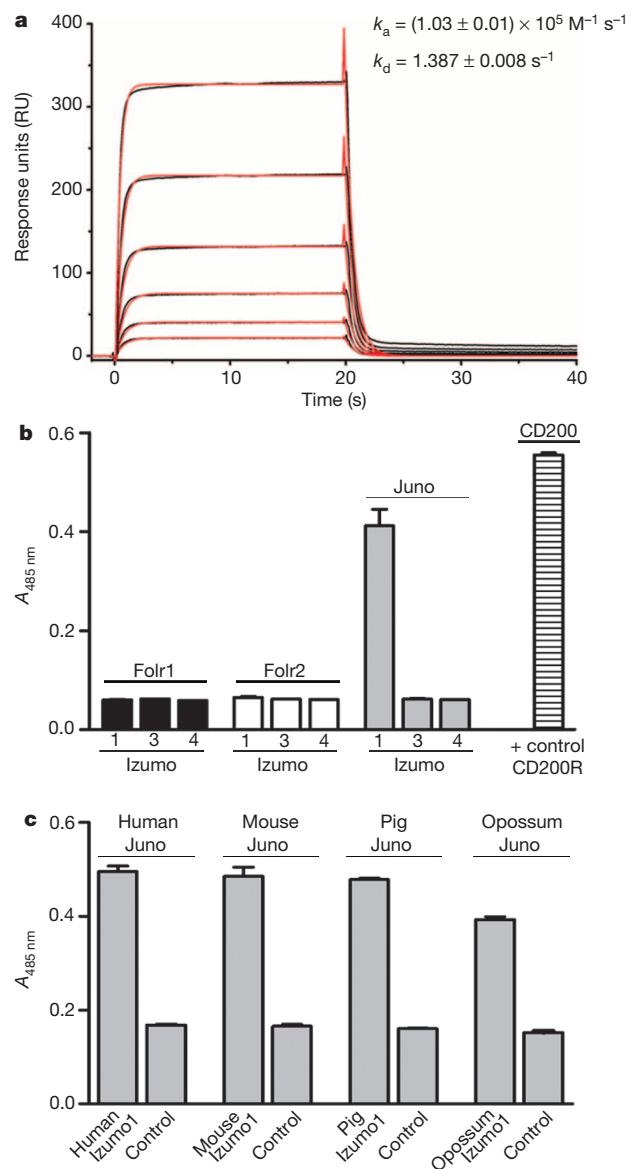
receptor 4 (*Folr4*) gene. *Folr4* is one of three folate receptor paralogues in mouse whose main role is thought to involve folate uptake. *Folr4* is expressed on CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells in mice<sup>13</sup> and is being tested as an anti-tumour therapy<sup>13–15</sup> as well as a mediator of responses to dietary folate<sup>16–18</sup>, but no functional role on oocytes has been reported. Using recombinant proteins, we showed that, unlike *Folr1* and *Folr2*, *Folr4* was unable to bind to folate (Extended Data Fig. 3a), which was expected given differences in amino acids known to be critical for folate binding<sup>19,20</sup> (Extended Data Fig. 3b). On the basis of these and subsequent findings, we renamed the gene encoding this protein 'Juno'. Using an anti-Juno monoclonal antibody, we showed that Juno expression matched the binding pattern of the recombinant Izumo probe on ovulated oocytes (Fig. 1c). Preincubating oocytes with the anti-Juno antibody prevented all detectable binding of the Izumo1 probe, demonstrating that Juno is the sole Izumo1 receptor on oocytes (Fig. 1d). Similar to other folate receptor paralogues, the protein sequence of Juno suggested the presence of a carboxy-terminal GPI-anchor site. Given the known importance of oolemmal GPI-linked proteins in fertilization<sup>10,11</sup>, we tested whether Juno was tethered to the membrane by a GPI anchor. A large fraction of cell-surface Juno staining was lost after phosphoinositide phospholipase C (PIPLC) treatment of either HEK293 cells transfected with the *Juno* cDNA (Fig. 1e) or oocytes (Fig. 1f), demonstrating that Juno was GPI-anchored, consistent with previous results<sup>21</sup>. Taken together, these data identify Juno as the Izumo1 receptor expressed on oocytes.

### Izumo1–Juno interaction is conserved

To show that Izumo1 and Juno interacted directly, and to quantify the biophysical parameters of the interaction, we expressed and purified the entire ectodomain of Juno and measured its binding to Izumo1 using surface plasmon resonance (SPR) (Fig. 2a). Like other extracellular interactions measured using this technique, the interaction was shown to be highly transient with a  $K_D$  of  $\sim 12 \mu\text{M}$  (Extended Data Fig. 4). Juno contains a single globular domain<sup>19,20</sup>, whereas Izumo1 contains both an immunoglobulin superfamily domain and a unique amino terminus termed the Izumo domain<sup>22</sup>. Consistent with the recent finding that the Izumo domain is sufficient for egg binding<sup>9</sup>, we expressed the individual domains of Izumo1 and showed, using an assay designed to detect direct low-affinity extracellular protein interactions called avidity-based extracellular interaction screen (AVEXIS)<sup>23</sup>, that it is the Izumo domain that contains both the OBF13 epitope and Juno binding site (Extended Data Fig. 5). Both Izumo1 and Juno belong to small paralogous gene families which, in mouse, contain four and three members, respectively. We expressed the entire ectodomains of Izumo1, 3 and 4 (the Izumo2 paralogue did not express) and used the AVEXIS assay to screen systematically for all possible pairwise interactions with the three *Folr* paralogues. We observed that only mouse Izumo1 and Juno could interact (Fig. 2b). Clearly identifiable *Izumo1* and *Juno* orthologues exist in all sequenced mammalian genomes, including marsupials. To determine whether the interaction was conserved within mammals, we expressed the entire ectodomains of both Izumo1 and Juno orthologues from humans, pig (*Sus scrofa*) and opossum (*Monodelphis domestica*) and assessed binding using the AVEXIS assay. Clear binding between the orthologues was observed, demonstrating that the interaction is conserved within mammals (Fig. 2c).

### Juno is essential for fertilization

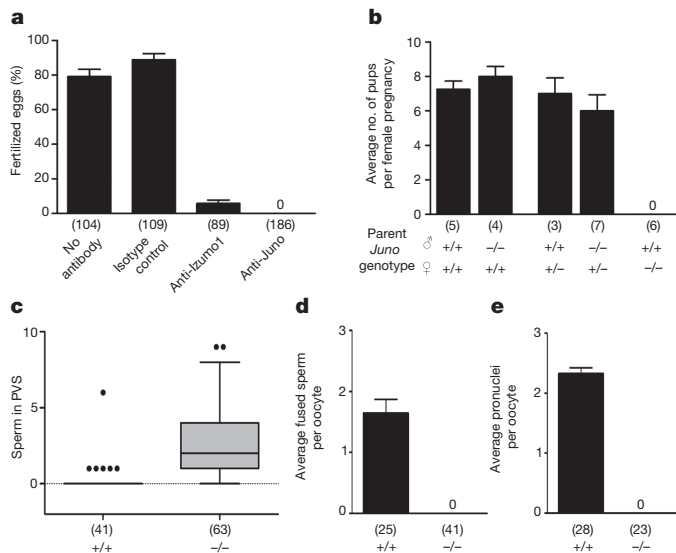
To assess the role of Juno in fertilization, we first added an anti-Juno monoclonal antibody that blocks the Izumo–Juno interaction (Fig. 1d) into *in vitro* fertilization (IVF) assays. The anti-Juno antibody potently prevented fertilization (Fig. 3a), with no detectable fertilization events even when used at concentrations as low as  $0.1 \mu\text{g ml}^{-1}$  (Extended Data Fig. 6). Using similar approaches, previous research in this area has identified promising candidates for sperm–egg recognition that, when genetically deleted, have been shown to be dispensable for fertilization<sup>3</sup>. To determine whether *Juno* was essential for fertilization, we created *Juno*-deficient mice using embryonic stem cells disrupted at the *Folr4*



**Figure 2 | The Izumo1–Juno interaction is direct, transient and conserved across mammals.** **a**, Biophysical analysis of the Izumo1–Juno interaction using SPR. Serial dilutions of purified, soluble Juno were injected over immobilised Izumo1, and kinetic parameters derived from a 1:1 Langmuir binding model (red lines). **b**, Binding specificity within the paralogues of the mouse Izumo and *Folr* families: only Juno (*Folr4*) bound Izumo1. **c**, The Izumo1–Juno interaction is conserved across mammals. AVEXIS was used for binding analysis in **b** and **c** using recombinant *Folr*/Juno proteins as preys and Izumo proteins as baits; control bait in **c** was CD4. All bar charts show mean  $\pm$  s.e.m.;  $n = 3$ .

locus using a gene trap allele: *Folr4*<sup>tm1a(KOMP)Wtsi</sup> (Extended Data Fig. 7). Both *Juno*<sup>−/−</sup> male and female mice developed indistinguishably from wild-type controls, showing normal rates of growth and morphology and were overtly healthy. Whereas *Juno*<sup>−/−</sup> male and *Juno*<sup>+/-</sup> female mice were fertile, by contrast, *Juno*<sup>−/−</sup> females failed to produce any litters during three months of continuous mating with wild-type males of proven fertility (Fig. 3b). Female *Juno*<sup>−/−</sup> mice exhibited natural mating behaviours, as assessed by vaginal plug formation and the presence of motile sperm in the reproductive tract when paired with fertile wild-type males. *Juno*<sup>−/−</sup> females responded to hormone treatment by ovulating morphologically normal eggs at numbers that did not significantly differ from wild-type. *Juno*<sup>−/−</sup> eggs recovered at embryonic day 0.5 after superovulation and natural mating were not fertilized and had more sperm within their perivitelline space compared to wild-type eggs, demonstrating that the zona pellucida of *Juno*<sup>−/−</sup> eggs could be



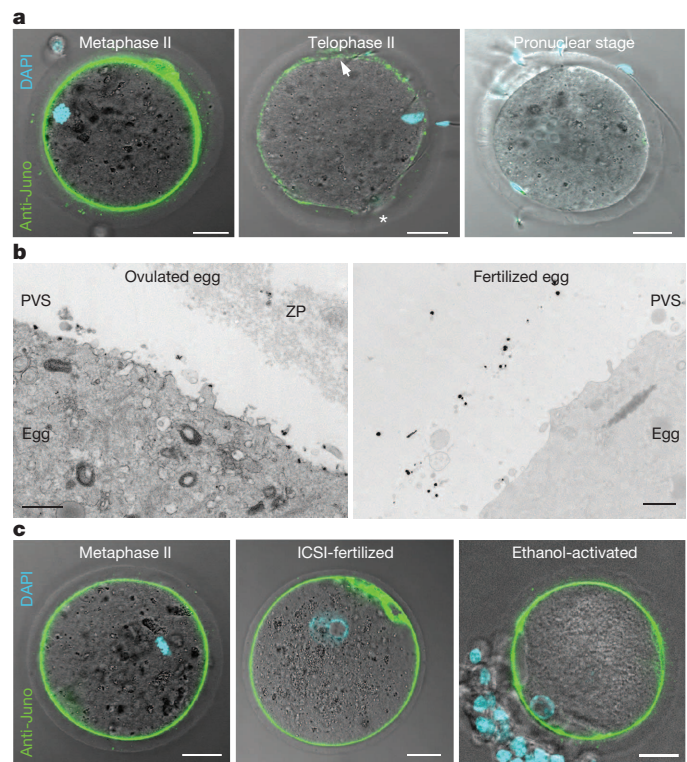


**Figure 3 | Juno is essential for female fertility.** **a**, An anti-Juno monoclonal antibody potentially blocked *in vitro* fertilization; anti-Izumo1 is shown for comparison. **b**, Female *Juno*<sup>-/-</sup> mice are infertile. Continuous mating of *Juno*<sup>-/-</sup> female mice to proven wild-type males for 3 months did not result in any pups. **c**, Greater numbers of sperm were observed in the perivitelline space (PVS) of eggs from superovulated *Juno*-deficient (-/-) mice relative to wild-type (+/+) mice. **d**, **e**, Zona-free *Juno*<sup>-/-</sup> eggs do not fuse with wild-type sperm *in vitro*. Sperm were added to *Juno*-deficient (-/-) and wild-type (+/+) eggs and fused sperm quantified after two hours (**d**), or pronuclei after six hours (**e**). All bar charts show mean  $\pm$  s.e.m.; numbers in parentheses are total number of eggs (**a**, **c**, **d**, **e**) or number of mating pairs (**b**).

penetrated by wild-type sperm *in vivo*, which then either did not bind or fuse with the oolemma (Fig. 3c). To investigate this further, we removed the zona pellucida from *Juno*<sup>-/-</sup> eggs and found that they could not be fertilized with wild-type sperm in IVF assays, as assessed by counting the number of fused sperm and formed pronuclei per egg (Fig. 3d, e). This is consistent with a role for the Izumo1–Juno interaction in sperm–egg adhesion or fusion. To distinguish between these possibilities, normally non-fusing HEK293 cells were separately transfected with plasmids encoding either *Juno* or *Izumo1* cDNAs and mixed. No syncytia were observed in the mixed cultures, indicating that the interaction is not sufficient for cell fusion (Extended Data Fig. 8a). Furthermore, acrosome-reacted normal sperm did not bind as efficiently to zona-free *Juno*-deficient eggs as wild-type controls (Extended Data Fig. 8b). Together, these findings indicate that the Izumo1–Juno interaction is a necessary adhesion event between sperm and egg that is required for fertilization.

### Juno is rapidly shed after fertilization

After fertilization, oocytes become largely refractory to further sperm fusion events to prevent the creation of nonviable polyploid embryos due to polyspermy. This is achieved through both a relatively slow-acting (>1 h) hardening of the zona pellucida caused by the action of enzymes released from cortical granules after egg activation, and a faster-acting block involving biochemical changes in the oolemmal membrane. In broadcast-spawning frogs and aquatic invertebrates, the membrane block to polyspermy is very fast (a few seconds to minutes) and is achieved by electrically depolarizing the oolemma<sup>24</sup>. In mammals, the membrane block occurs over a longer timeframe (30–45 min), and despite being first described 60 years ago, its mechanistic basis remains a mystery<sup>2</sup>. Given the essential role of Juno in fertilization, we explored the role it might play in establishing the membrane block to polyspermy. Interestingly, cell-surface Juno was rapidly lost after fertilization, being only weakly detectable in zona-intact fertilized eggs at telophase II and undetectable at the pronuclear stage (Fig. 4a). Similarly, Juno was barely detectable on anaphase-II-stage zona-free oocytes, approximately 30 to 40 min after fertilization (Extended Data Fig. 9). Folate receptors are



**Figure 4 | Juno is rapidly shed from the oolemma of normally fertilized eggs, but not ICSI-fertilized or parthenogenetically activated eggs.**

**a**, Cell-surface Juno rapidly becomes undetectable after fertilization. Juno (green) is expressed on ovulated metaphase II eggs, but is barely evident at telophase II and undetectable on pronuclear-stage zygotes. Arrow and asterisk indicate sites of first and second polar body extrusion, respectively; chromosomes are not within the plane of focus. **b**, Immunogold electron microscopy localized Juno primarily to the oolemma in unfertilized, ovulated eggs, but was redistributed to vesicles within the perivitelline space within 1 h after fertilization. ZP, zona pellucida. **c**, Eggs fertilized by ICSI or parthenogenetically activated retain oolemmal Juno staining until at least the pronuclear stage, as shown. Nuclei/pronuclei were visualized with DAPI (blue). Images show representative eggs from three independent experiments. Scale bars represent 20  $\mu$ m (**a**, **c**) and 1  $\mu$ m (**b**).

known to take up extracellular dietary folate by endocytosis<sup>25</sup> and we therefore used immunogold electron microscopy to determine the sub-cellular localization of Juno following fertilization. Surprisingly, we observed that Juno was not internalised after fertilization but present in extracellular vesicles, presumably derived from the microvillus-rich oolemma that undergoes significant architectural changes upon fertilization<sup>26</sup> (Fig. 4b). Some methods of fertilization, such as intracytoplasmic sperm injection (ICSI) and parthenogenetic activation, do not lead to the establishment of an effective membrane polyspermic block<sup>2</sup>. Consistent with a role for Juno in the membrane block to polyspermy, eggs fertilized either by ICSI, or parthenogenetically activated by ethanol, did not lose cell-surface expression of Juno (Fig. 4c). The rapid shedding of Juno from the egg membrane within vesicles after fertilization therefore provides a possible mechanism for the membrane block to polyspermy in mammalian eggs.

### Discussion

To our knowledge, the discovery of Juno as the binding partner for Izumo1 identifies these proteins as the first cell-surface receptor pair essential for gamete recognition in any organism. Importantly, both female *Juno* and male *Izumo1*-deficient mice are infertile, suggesting that the interaction is essential for normal fertilization. Comparing our findings with recent advances in the molecular understanding of cellular fusion in other biological contexts suggests that the Izumo1–Juno interaction performs a necessary adhesion step rather than acting as a

membrane 'fusogen'. Unlike established fusogens such as the EFF-1 and AFF-1 proteins<sup>27</sup> or syncytins<sup>28</sup>, that are sufficient for fusion, Juno and Izumo1 do not induce the formation of syncytia when ectopically expressed on neighbouring non-fusing cells. Essential adhesion processes are required in other cellular fusion systems such as myofibre formation in both *Drosophila*<sup>29</sup> and vertebrates<sup>30</sup>, suggesting that membrane fusion requires the action of other membrane proteins such as myomaker<sup>31</sup>. The expected but remarkably low *in vitro* binding affinity between soluble monomeric Juno and Izumo1 suggests that local clustering of Juno within the oolemma is important to ensure a sufficiently high binding avidity for productive interactions, a role that could be performed by the tetraspanin CD9. Tetraspanins are known to organize other membrane proteins within microdomains<sup>32</sup>, and quantitative differences in sperm adhesion to CD9-deficient eggs have been reported<sup>33</sup>. This may also explain why female mice lacking CD9 have reduced fertility<sup>5–7</sup>, unlike Juno-deficient mice which are completely infertile. The ability of CD9 to organize Juno within the oolemma may be indirect because oocytes lacking CD9 have an altered membrane architecture<sup>34,35</sup>.

Several pieces of evidence point to the rapid loss of Juno from the egg membrane after fertilization as the mechanistic basis for the membrane block to polyspermy in mammals. First, cell-surface Juno is essential for fertilization; second, it is undetectable on the oolemma approximately 40 min after fertilization, in excellent agreement with the timing of the membrane block<sup>36</sup>; third, the membrane block is both a graded response<sup>37</sup> and associated with the loss of sperm binding sites<sup>38</sup>, consistent with the gradual loss of a surface receptor; and fourth, surface Juno expression is not lost in ICSI-fertilized or parthenogenetically-activated eggs that do not form an effective membrane block<sup>39</sup>. One further possibility is that because Juno is shed as vesicles after fertilization, this creates a zone of 'decoy eggs' confined within the perivitelline space that could bind to and efficiently neutralise incoming acrosome-reacted sperm to increase the potency of the sperm block. In conclusion, our discovery of an extracellular receptor pair essential for fertilization provides a focus for the rational development of novel contraceptives and fertility treatments.

## METHODS SUMMARY

**Transgenic mice.** All animal experiments were performed in accordance with local ethical and UK Home Office regulations. Juno-deficient mice (*Folr4*<sup>tm1a(KOMP)Wtsl</sup>) were generated by the Sanger Institute Mouse Genetics Project from *Folr4*-targeted mouse embryonic stem cells as part of the International Knockout Mouse Consortium, essentially as described<sup>40</sup>. Acrosome reporter mice (B6;B6C3-Tg(Acro3-EGFP)010sb)<sup>41</sup>, where the enhanced green fluorescent protein labels acrosome-intact sperm, were obtained from the RIKEN BioResource Centre. Wild-type eggs were obtained from B/CB strain mice for IVF, parthenogenetic activation and ICSI procedures. Sibling C57BL/6N wild-types were used as controls in all other experiments. **In vitro fertilization assays and ICSI.** IVF and ICSI were performed as previously described<sup>42,43</sup>. Eggs were collected from hormone-primed females and sperm were capacitated before mixing. Fertilization and egg activation were assessed by staining sperm DNA and egg chromosomes. Acrosome-reacted sperm were identified using Acro3-EGFP reporter mice. Juno expression analysis was performed by immunofluorescence and images acquired with a Leica SP5 confocal microscope. Fertilization was quantified by counting decondensed sperm heads in the egg cytoplasm, or scoring the number of pronuclei.

**Recombinant protein production and expression cloning.** All protein expression constructs were made by gene synthesis (GeneArt). Protein production, purification, AVExis assays and SPR were performed essentially as described<sup>23,44</sup>. The Juno cDNA was identified by iterative rounds of selection by transfecting adherent HEK293T cells with a commercially available mouse oocyte cDNA library, NIH\_MGC\_257\_N (Express Genomics, USA).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 January; accepted 5 March 2014.

Published online 16 April 2014.

- Okabe, M. The cell biology of mammalian fertilization. *Development* **140**, 4471–4479 (2013).

- Gardner, A. J. & Evans, J. P. Mammalian membrane block to polyspermy: new insights into how mammalian eggs prevent fertilisation by multiple sperm. *Reprod. Fertil. Dev.* **18**, 53–61 (2006).
- Ikawa, M., Inoue, N., Benham, A. M. & Okabe, M. Fertilization: a sperm's journey to and interaction with the oocyte. *J. Clin. Invest.* **120**, 984–994 (2010).
- Inoue, N., Ikawa, M., Isotani, A. & Okabe, M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* **434**, 234–238 (2005).
- Le Naour, F., Rubinstein, E., Jasmin, C., Prenant, M. & Boucheix, C. Severely reduced female fertility in CD9-deficient mice. *Science* **287**, 319–321 (2000).
- Miyado, K. *et al.* Requirement of CD9 on the egg plasma membrane for fertilization. *Science* **287**, 321–324 (2000).
- Kaji, K. *et al.* The gamete fusion process is defective in eggs of Cd9-deficient mice. *Nature Genet.* **24**, 279–282 (2000).
- Satouh, Y., Inoue, N., Ikawa, M. & Okabe, M. Visualization of the moment of mouse sperm-egg fusion and dynamic localization of IZUMO1. *J. Cell Sci.* **125**, 4985–4990 (2012).
- Inoue, N. *et al.* Molecular dissection of IZUMO1, a sperm protein essential for sperm-egg fusion. *Development* **140**, 3221–3229 (2013).
- Coonrod, S. A. *et al.* Treatment of mouse oocytes with PI-PLC releases 70-kDa (pI 5) and 35- to 45-kDa (pI 5.5) protein clusters from the egg surface and inhibits sperm-oolemma binding and fusion. *Dev. Biol.* **207**, 334–349 (1999).
- Alfieri, J. A. *et al.* Infertility in female mice with an oocyte-specific knockout of GPI-anchored proteins. *J. Cell Sci.* **116**, 2149–2155 (2003).
- Wright, G. J. Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. *Mol. Biosyst.* **5**, 1405–1412 (2009).
- Yamaguchi, T. *et al.* Control of immune responses by antigen-specific regulatory T cells expressing the folate receptor. *Immunity* **27**, 145–159 (2007).
- Teng, M. W. *et al.* Multiple antitumor mechanisms downstream of prophylactic regulatory T-cell depletion. *Cancer Res.* **70**, 2665–2674 (2010).
- Liang, S. C., Moskalenko, M., Van Roey, M. & Jooss, K. Depletion of regulatory T cells by targeting folate receptor 4 enhances the potency of a GM-CSF-secreting tumor cell immunotherapy. *Clin. Immunol.* **148**, 287–298 (2013).
- Kunisawa, J., Hashimoto, E., Ishikawa, I. & Kiyono, H. A pivotal role of vitamin B9 in the maintenance of regulatory T cells *in vitro* and *in vivo*. *PLoS ONE* **7**, e32094 (2012).
- Kinoshita, M. *et al.* Dietary folic acid promotes survival of Foxp3<sup>+</sup> regulatory T cells in the colon. *J. Immunol.* **189**, 2869–2878 (2012).
- Salbaum, J. M., Kruger, C. & Kappen, C. Mutation at the folate receptor 4 locus modulates gene expression profiles in the mouse uterus in response to periconceptional folate supplementation. *Biochim. Biophys. Acta* **1832**, 1653–1661 (2013).
- Chen, C. *et al.* Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **500**, 486–489 (2013).
- Wibowo, A. S. *et al.* Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. *Proc. Natl Acad. Sci. USA* **110**, 15180–15188 (2013).
- Jia, Z. *et al.* A novel splice variant of FR4 predominantly expressed in CD4<sup>+</sup>CD25<sup>+</sup> regulatory T cells. *Immunol. Invest.* **38**, 718–729 (2009).
- Ellerman, D. A. *et al.* Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol. Reprod. Dev.* **76**, 1188–1199 (2009).
- Bushell, K. M., Sollner, C., Schuster-Boeckler, B., Bateman, A. & Wright, G. J. Large-scale screening for novel low-affinity extracellular protein interactions. *Genome Res.* **18**, 622–630 (2008).
- Jaffe, L. A. Fast block to polyspermy in sea urchin eggs is electrically mediated. *Nature* **261**, 68–71 (1976).
- Sabharanjak, S. & Mayor, S. Folate receptor endocytosis and trafficking. *Adv. Drug Deliv. Rev.* **56**, 1099–1109 (2004).
- Jackowski, S. & Dumont, J. N. Surface alterations of the mouse zona pellucida and ovum following *in vivo* fertilization: correlation with the cell cycle. *Biol. Reprod.* **20**, 150–161 (1979).
- Podbilewicz, B. *et al.* The *C. elegans* developmental fusogen EFF-1 mediates homotypic fusion in heterologous cells and *in vivo*. *Dev. Cell* **11**, 471–481 (2006).
- Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–789 (2000).
- Abmayr, S. M. & Pavlath, G. K. Myoblast fusion: lessons from flies and mice. *Development* **139**, 641–656 (2012).
- Powell, G. T. & Wright, G. J. Jamb and Jamc are essential for vertebrate myocyte fusion. *PLoS Biol.* **9**, e1001216 (2011).
- Millay, D. P. *et al.* Myomaker is a membrane activator of myoblast fusion and muscle formation. *Nature* **499**, 301–305 (2013).
- Hemler, M. E. Tetraspanin functions and associated microdomains. *Nature Rev. Mol. Cell Biol.* **6**, 801–811 (2005).
- Jégou, A. *et al.* CD9 tetraspanin generates fusion competent sites on the egg membrane for mammalian fertilization. *Proc. Natl Acad. Sci. USA* **108**, 10946–10951 (2011).
- Runge, K. E. *et al.* Oocyte CD9 is enriched on the microvillar membrane and required for normal microvillar shape and distribution. *Dev. Biol.* **304**, 317–325 (2007).
- Zylkiewicz, E., Nowakowska, J. & Maleszewski, M. Decrease in CD9 content and reorganization of microvilli may contribute to the oolemma block to sperm penetration during fertilization of mouse oocyte. *Zygote* **18**, 195–201 (2010).
- Wolf, D. P. The block to sperm penetration in zona-free mouse eggs. *Dev. Biol.* **64**, 1–10 (1978).
- Gardner, A. J., Williams, C. J. & Evans, J. P. Establishment of the mammalian membrane block to polyspermy: evidence for calcium-dependent and -independent regulation. *Reproduction* **133**, 383–393 (2007).



38. Horvath, P. M., Kellom, T., Caulfield, J. & Boldt, J. Mechanistic studies of the plasma membrane block to polyspermy in mouse eggs. *Mol. Reprod. Dev.* **34**, 65–72 (1993).
39. Wortzman-Show, G. B., Kurokawa, M., Fissore, R. A. & Evans, J. P. Calcium and sperm components in the establishment of the membrane block to polyspermy: studies of ICSI and activation with sperm factor. *Mol. Hum. Reprod.* **13**, 557–565 (2007).
40. Skarnes, W. C. *et al.* A conditional knockout resource for the genome-wide study of mouse gene function. *Nature* **474**, 337–342 (2011).
41. Nakanishi, T. *et al.* Real-time observation of acrosomal dispersal from mouse sperm using GFP as a marker protein. *FEBS Lett.* **449**, 277–283 (1999).
42. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo: A Laboratory Manual* 3rd edn (Cold Spring Harbor Laboratory Press, 2003).
43. Kimura, Y. & Yanagimachi, R. Intracytoplasmic sperm injection in the mouse. *Biol. Reprod.* **52**, 709–720 (1995).
44. Sun, Y., Gallagher-Jones, M., Barker, C. & Wright, G. J. A benchmarked protein microarray-based platform for the identification of novel low-affinity extracellular protein interactions. *Anal. Biochem.* **424**, 45–53 (2012).

**Acknowledgements** This work was supported by the Wellcome Trust grant number 098051. *Juno*-deficient mice were generated by the Sanger Institute Mouse Genetics Project. We thank W. Skarnes and J. Bussell for advice on transgenic mice; J. Kerr for construct design; A. Bradley and L. Jovine for helpful comments on the manuscript; and M. Okabe and N. Inoue for the OBF13 antibody.

**Author Contributions** E.B. and G.J.W. conceived the project, designed and analysed the experiments, and wrote the manuscript. E.B. performed all experiments with technical help from B.D. (ICSI) and D.G. (electron microscopy).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.J.W. ([gw2@sanger.ac.uk](mailto:gw2@sanger.ac.uk)).

# Origins and functional evolution of Y chromosomes across mammals

Diego Cortez<sup>1,2</sup>, Ray Marin<sup>1,2</sup>, Deborah Toledo-Flores<sup>3</sup>, Laure Froidevaux<sup>1</sup>, Angélica Liechti<sup>1</sup>, Paul D. Waters<sup>4</sup>, Frank Grützner<sup>3</sup> & Henrik Kaessmann<sup>1,2</sup>

**Y chromosomes underlie sex determination in mammals, but their repeat-rich nature has hampered sequencing and associated evolutionary studies. Here we trace Y evolution across 15 representative mammals on the basis of high-throughput genome and transcriptome sequencing. We uncover three independent sex chromosome originations in mammals and birds (the outgroup). The original placental and marsupial (therian) Y, containing the sex-determining gene *SRY*, emerged in the therian ancestor approximately 180 million years ago, in parallel with the first of five monotreme Y chromosomes, carrying the probable sex-determining gene *AMH*. The avian W chromosome arose approximately 140 million years ago in the bird ancestor. The small Y/W gene repertoires, enriched in regulatory functions, were rapidly defined following stratification (recombination arrest) and erosion events and have remained considerably stable. Despite expression decreases in therians, Y/W genes show notable conservation of proto-sex chromosome expression patterns, although various Y genes evolved testis-specificities through differential regulatory decay. Thus, although some genes evolved novel functions through spatial/temporal expression shifts, most Y genes probably endured, at least initially, because of dosage constraints.**

In most mammals, Y chromosomes are required to override the program underlying development of the default sex, females<sup>1</sup>. Extant mammals possess an XY (male heterogametic) sex chromosome system, with rare exceptions that experienced secondary XY loss<sup>2</sup>, but sex chromosomes evolved from different autosomes in therians (placentals and marsupials) and egg-laying monotremes (Fig. 1). Therians share the same XY system, whereas monotremes have multiple X and Y chromosomes that are partially homologous to bird sex chromosomes<sup>3–5</sup>, where females are heterogametic (ZW system). Sex chromosome differentiation occurred through recombination arrests along the Y, leading to reduced selection and associated gene decay and repeat accumulation<sup>6,7</sup>. Consequences of Y deterioration for X chromosome evolution were recently investigated<sup>8</sup>, but the fact that this chromosome is refractory to assembly owing to its repeat-rich nature<sup>7</sup> have hindered evolutionary studies of the Y itself. Nevertheless, dedicated efforts determined complete Y sequences in three primates<sup>9–11</sup> (human, chimpanzee, macaque) and large portions of two carnivore (dog and cat) Y chromosomes<sup>12</sup>. Together with smaller scale work, these studies provided initial clues to Y evolution, such as the stabilization of Y gene content through ampliconic sequence accumulation<sup>10,11,13</sup>. However, our understanding of mammalian Y chromosome evolution remains limited owing to the restricted amount and phylogenetic representation of available Y chromosome data.

## Mammalian Y gene repertoires

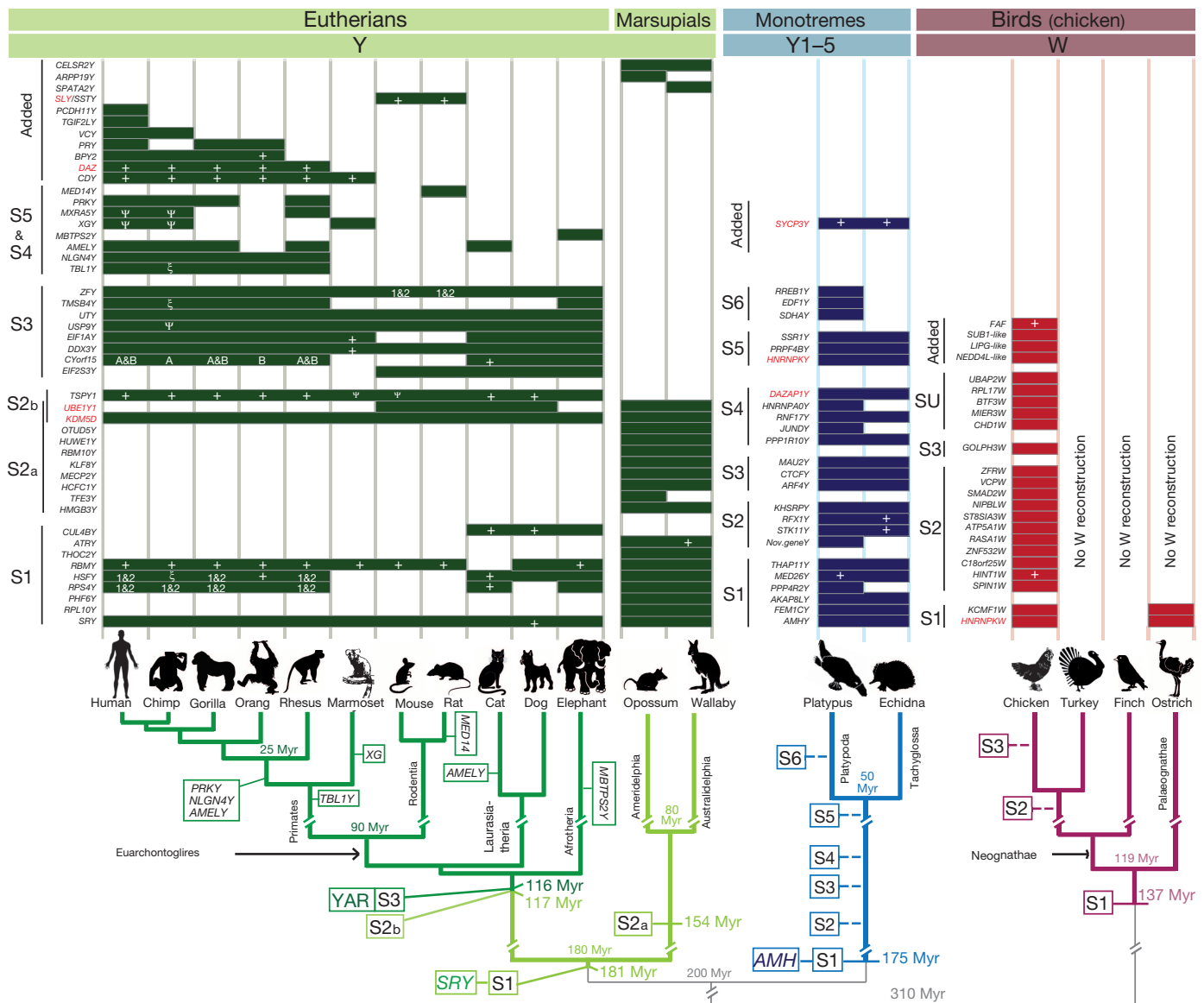
To explore Y evolution, we developed a subtraction approach that directly targets and assembles exons of the male-specific (non-pseudoautosomal) Y chromosome (MSY) on the basis of high-throughput sequencing of transcriptomes and genomes from both sexes (Extended Data Fig. 1 and Methods). In brief, male transcripts were assembled from male-specific RNA sequencing (RNA-seq) reads not mapping onto female reference genomes. Y identity was confirmed using the whole-genome sequence data; that is, true Y transcripts are supported by genomic reads unique to males. Genes with no or low expression were detected by screening

male-specific genomic data with Y orthologues from other species. The genomic data also served to support the absence of ancestral Y genes (that is, their evolutionary loss). We validated our approach using large-scale PCR/Sanger sequencing-based screening of male/female genomic DNA and published Y chromosomes, thus confirming that complete coding sequences of all Y genes (with the potential exception of those not expressed in the sampled tissues and lacking a known Y orthologue) could be deduced for a given species.

We applied our procedure to sequencing data that we collected for ten mammals (Supplementary Tables 1–4 and Methods). Together with available Y sequences<sup>9–12</sup>, we could thus investigate Y evolution in 15 species covering all major lineages of the class Mammalia<sup>14</sup> (placentals or eutherians, marsupials, monotremes), all but one (Xenarthra) of the eutherian superorders<sup>14</sup>, the two marsupial superorders (American and Australian marsupials), both extant monotreme families (platypus, echidna), and all major groups of ‘higher primates’ (that is, simians: the great apes, Old World monkeys (OWMs) and New World monkeys (NWMs)). For comparison, we produced similar data for the bird (chicken) W chromosome.

We identified a total of 134 different Y protein-coding genes in the ten new species, thus approximately doubling the number of previously known Y genes (Fig. 1 and Extended Data Fig. 3; note that the letter Y is added at the end of gene symbols for all genes for which Y-linkage is detected, following common practice), and 214 distinct pseudogenes and noncoding RNAs (Supplementary Tables 5–17 and Supplementary Data 1). Our read coverage analysis (Methods) predicts 0–6 multi-copy protein-coding genes per species, some of which are shared across species and thus presumably belong to conserved ampliconic Y regions<sup>11</sup> (Fig. 1 and Supplementary Tables 5–17). Most of the 155 distinct noncoding RNAs seem to stem from ampliconic regions, with two or more identical copies per locus (Supplementary Tables 5–17). We recovered all of the recently identified chicken W protein-coding genes<sup>15,16</sup> and added 11 distinct noncoding RNAs (Fig. 1 and Supplementary Table 18).

<sup>1</sup>Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. <sup>2</sup>Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland. <sup>3</sup>The Robinson Research Institute, School of Molecular and Biomedical Science, University of Adelaide, Adelaide, South Australia 5005, Australia. <sup>4</sup>School of Biotechnology and Biomolecular Sciences, UNSW Australia, Sydney, New South Wales 2052, Australia.



**Figure 1 | Y (W) protein-coding gene repertoires and their origins.** Top, Y/W genes and strata (genes are ordered within strata based on  $d_s$ ). Genes added to differentiated sex chromosomes ('Added') and independently recruited genes (red text) are indicated. For echidna and ostrich, only genes shared with sister species (platypus and chicken, respectively) were identified on the basis of genomic sequencing data. For bird strata age estimations, turkey and zebra finch Z gametologues were included. Bottom, phylogenetic tree of the 19 species. Lineages, divergence times and origins of strata and genes are indicated. For therian strata, monotreme S1 and bird S1, refined age estimates

### Origins of therian Y genes

We next assessed the time at which Y genes began to evolve independently of their X counterpart (gametologue), or when they were recruited (through transposition/translocation) to already differentiated Y regions from autosomes or the X. We thus reconstructed phylogenetic trees for all Y genes, X gametologues, and autosomal orthologues from outgroup species with different sex chromosomes (Extended Data Fig. 4, Supplementary Data 1 and Methods). In these trees, the location of nodes reflecting the divergence of Y and X clades, relative to lineage divergence nodes, mark the phylogenetic position at which a given Y gene emerged. We assessed divergence rates at synonymous sites ( $d_s$ ) before/after differentiation of Y/X gametologues for further refinement of ages (Extended Data Fig. 5a–f).

Y chromosomes consist of distinct regions ('strata') that stopped recombining at specific time points<sup>10,17</sup>. Our phylogenetic dating of the

on branches are based on  $d_s$  trees (Extended Data Fig. 5a–f). More approximate placement on branches for platypus S2–S5 and chicken S2–S3 (genes for which strata could not be assigned with confidence are listed as undefined, SU) are based on pairwise (median)  $d_s$  values of stratum genes (dashed lines). Note that although actual Y chromosome assemblies underlie the previously published<sup>12</sup> dog/cat Y repertoires shown here, the repertoires may be incomplete, because these assemblies (in particular for cat) are incomplete<sup>12</sup>. +, multi-copy genes;  $\psi$ , pseudogenes;  $\xi$ , intact genes in chimpanzee formerly annotated as pseudogenes; A&B 1&2, known duplicates; YAR, Y-added region.

original eutherian stratum 1 (S1), which, like S2, stems from the same ancestral proto-sex chromosome portion (the 'Y/X-conserved' region, YCR/XCR; ref. 18), shows that it originated in the last common therian ancestor (Fig. 1 and Extended Data Fig. 4a; see the following external link for individual trees: [ftp://ftp.vital-it.ch/papers/kaessmann/Nature-Cortez/Cortez\\_etal\\_Nature\\_YX\\_gene\\_trees\\_alignments.zip](http://ftp.vital-it.ch/papers/kaessmann/Nature-Cortez/Cortez_etal_Nature_YX_gene_trees_alignments.zip)), confirming that sex chromosomes of humans and other therians emerged after the therian–monotreme split<sup>3,4</sup> and were never common to all mammals, as implied by some previous work<sup>10,19</sup>. Notably, therian sex chromosome origination occurred just before the eutherian–marsupial split ~180 million years (Myr) ago (Fig. 1 and Extended Data Fig. 5a), consistent with our previous hypothesis<sup>3,8</sup>. S1 contains a common set of four genes, including the sex-determining gene *SRY*<sup>1</sup>, which has been maintained since sex chromosome inception in most therians, although some S1 genes apparently became dispensable in some species (Fig. 1). Interestingly,



marsupials retained twice the number of ancestral S1 genes. Finally, it is noteworthy that  $d_s$  for Y gametologues substantially increased relative to that of X counterparts in therians following S1 differentiation (Benjamini–Hochberg-corrected  $P < 0.05$ , Mann–Whitney  $U$ -test; Extended Data Figs 5g and 7a–d), suggesting that male mutation bias (that is, a higher mutation rate in males than in females owing to a higher number of germline cell divisions), previously reported for eutherians<sup>20</sup>, has shaped genomes of therians since their emergence.

S2 arose ~117 Myr ago in the eutherian ancestor, after the split from the marsupial lineage (Extended Data Figs 4b and 5b), consistent with some previous studies<sup>21</sup> but contrary to others<sup>19</sup>. In marsupials an independent S2, containing twice the number of eutherian S2 genes, arose in the marsupial ancestor ~37 Myr before the eutherian S2 (Extended Data Figs 4e and 5c). Thus, the therian pseudoautosomal region (PAR) shrank at different rates in eutherians and marsupials, involving the convergent differentiation of two genes (*KDM5D*, *UBE1Y1* (also known as *UBA1Y*)), as previously proposed<sup>21</sup> (Fig. 1).

Our analysis of S3, a large autosomal addition to the already established sex chromosomes<sup>18</sup> (the ‘Y/X-added region’), shows that it contained seven genes and was already defined in the common eutherian ancestor (Fig. 1 and Extended Data Fig. 4c), as could previously only be surmised<sup>10,17</sup>. S3 only differentiated just before the placental mammal radiation ~116 Myr ago, at about the same time or perhaps even concomitantly with S2 (for example, as part of the same inversion event; Extended Data Fig. 5d), implying that the eutherian MSY consisted merely of S1 for ~60 Myr following the eutherian–marsupial split. As for S1/S2, S3 gene content has been remarkably stable during evolution (Fig. 1). However, curiously, similarly to S1 (*HSFY*, *RPS4Y*), rodents and marmosets independently lost the same two S3 genes (*TMSB4Y*, *Cyorf15Y*), whereas *TMSB4Y* was independently also lost in Laurasiatheria<sup>22</sup> (Fig. 1). It is also noteworthy that the elephant retained six more ancestral S1–S3 genes (total 15) than the marmoset (total nine), the eutherian with the smallest detected S1–S3 gene repertoire.

Further analyses show that all genes of S4 and S5, as previously defined<sup>10</sup>, originated in the common catarrhine (OWM/great ape) ancestor ~25–40 Myr ago, except for one gene (*TBL1Y*), which presumably emerged in the common simian ancestor >40 Myr ago (Fig. 1). The marmoset Y acquired a unique gene, *XG* (Fig. 1), which spans the human PAR boundary<sup>23</sup> and remains in the elephant PAR region<sup>24</sup>. Furthermore, *AMELY*, also residing near the PAR boundary, probably emerged independently in primates and Laurasiatheria<sup>21</sup>. *MBTPS2Y* arose as part of an independent S4 in the afrotherian lineage, and *MED14Y* represents a recent acquisition in rats (Fig. 1). Primate Y chromosomes recruited additional genes through various (retro)transpositions/translocations from autosomes during recent evolution<sup>10,11</sup> (Fig. 1). These results illustrate the dynamic recent, partly convergent, evolution of the eutherian Y.

Although extant gene numbers are small, the above results illustrate that many Y genes endured over long evolutionary time periods. Modelling of Y gene loss dynamics revealed that S1–S3 gene decay proceeded rapidly upon differentiation, at similar rates as previously estimated<sup>10</sup>, and then markedly levelled off when small but apparently essential gene repertoires were defined (Extended Data Fig. 8c).

## Monotreme Y biology and evolution

Gene content and evolution of monotreme Y chromosomes (five in platypus, four in echidna owing to a  $Y_5$ – $Y_3$  fusion<sup>25</sup>) previously remained largely unknown, and the X chromosomes of platypus, the only monotreme with an assembled genome<sup>26</sup>, are only partly reconstructed. To assess monotreme sex chromosome evolution, we dated the 25 identified monotreme Y protein-coding genes (Fig. 1) and X gametologues. Because no X gametologue is assigned to any currently assembled X, we confirmed X identity for 23 out of 25 gametologues and assessed probable locations on the basis of male/female genomic read coverage and synteny mapping (Extended Data Fig. 6, Supplementary Tables 19, 20 and Methods), revealing that monotreme gametologues stem from various

ancestral synteny blocks (Supplementary Tables 19, 20) that assembled during evolution through genomic rearrangements<sup>4</sup>.

We detected six potential Y chromosome strata, five of which emerged in the platypus–echidna ancestor (Fig. 1 and Extended Data Fig. 5h). Thus, nearly all (22 out of 25) platypus Y genes differentiated at least ~50 Myr ago in the common monotreme ancestor (Supplementary Data 1, external data link). Importantly, monotreme S1 originated early during monotreme evolution ~175 Myr ago (Extended Data Figs 4f and 5e), which, together with our dating of therian sex chromosome origination, implies two independent yet essentially concomitant sex chromosome origination events in mammals, rules out the possibility of a turnover of one mammalian sex chromosome system into the other, and raises the question of the nature of sex determination in the common mammalian ancestor more than ~200 Myr ago.

Notably, one of the S1 genes is *AMH*, encoding the anti-Müllerian hormone, a key component of sex determination cascades across vertebrates<sup>1,27</sup>, which in eutherians blocks the development of female reproductive organs upon activation by *SOX9*, which in turn is activated by *SRY*<sup>1</sup>. Our observation that *AMH* is Y-linked in monotremes and part of the oldest stratum, S1 (Fig. 1 and Extended Data Fig. 5h), together with findings that *AMH* may precede *SOX9* expression and act as the primary sex-determining trigger in some vertebrates<sup>27–29</sup>, render this gene a prime candidate for being a principal monotreme sex-determining gene. We predicted *AMHX* to be located on chromosome  $X_1$  (Supplementary Tables 19, 20) and *AMHY* to be part of  $Y_5$ , on the basis of our synteny approach and the recent mapping<sup>30</sup> of the S1 gene *MED26Y* (Fig. 1), which similarly to *AMHY* stems from an ancestral linkage group corresponding to chicken chromosome 28. Notably, we confirmed that *AMHY* is located on  $Y_5$  by physical mapping (Extended Data Fig. 7e, f and Methods). Our results favour a scenario in which a fusion of an ancestral chromosome segment containing *AMH*, presumably to the original proto- $Y_5$ – $X_5$ , formed the initial monotreme proto-sex chromosomes. Their differentiation was followed by various translocations/fusions with autosomes<sup>25</sup> and subsequent differentiation events, which ultimately led to the current Y genes, of which only two genes (*FEM1CY* and *HNRNPKY*) were derived from the initial Z-homologous proto- $Y_5$  portion. Curiously, in parallel to eutherians, monotreme Y chromosomes also acquired *DAZ* and *SLY* homologous genes (Fig. 1).

Finally, contrary to our therian observations, we could not detect significantly different  $d_s$  values between monotreme Y and X gametologues (Extended Data Fig. 5g), which indicates that male mutation bias<sup>20</sup> is lacking or limited in monotremes.

## W origin

Next, we performed the first phylogenetic dating of chicken W strata (Fig. 1 and Supplementary Data 1, external data link), taking advantage of Z orthologues from recent bird genome sequences<sup>31,32</sup> and male/female RNA-seq data from ostriches<sup>33</sup> (allowing the extraction of ostrich W genes), representing the most basal avian lineage, ratites. Contrary to previous work<sup>34</sup>, this analysis reveals an initial stratum with two genes (*HNRNPK*, *KCMF1*) that emerged in the common bird ancestor ~140 Myr ago (Extended Data Figs 4g and 5f), whereas the remaining chicken W strata seem to have originated later during avian evolution (Fig. 1), although a more precise phylogenetic dating of these strata will require W sequences from other birds. *KCMF1* and *HNRNPK*, which was independently retained on both the W and monotreme Y (reflecting their shared ancestry; Fig. 1), are widely expressed housekeeping genes<sup>35,36</sup> that do not represent obvious sex-determination candidates but rather dosage-sensitive genes that were preserved to maintain proto-Z/W gene dose. Our results are consistent with a Z dosage-based mechanism of sex determination in birds<sup>15,37</sup>.

## Functional evolution of Y (W) genes

To understand why only small specific subsets of 9–25 Y/W protein-coding genes have been preserved from the original proto-sex chromosome repertoires, we assessed whether they possess characteristic

functions using simulations, which started with the ancestral sets of proto-sex-linked genes and then randomly removed genes until current Y/W gene numbers were reached (Methods). We then compared the functions of the simulated gene sets to those observed using Gene Ontology (GO) annotations<sup>38</sup>. Notably, this analysis revealed that highly non-random gene sets with similar functions were maintained across the different Y/W chromosomes (corrected  $P < 0.01$ , one-tailed alpha test). Current Y/W genes are enriched for genes involved in transcription/transcription regulation (GO Biological Process) and specific DNA binding/transcription factor activity (GO Molecular Function) (Supplementary Tables 21, 22), which is consistent with recent observations in fruitflies<sup>39</sup> and indicates that current Y genes were, at least initially, preserved to maintain ancestral gene dosage<sup>39</sup>, given that regulatory genes and genes with binding functions are often haploinsufficient<sup>40</sup>.

Expression levels of autosomal orthologues of Y/X genes in species with different sex chromosome systems, unaffected by sex-related selection, may serve to gauge proto-sex chromosome expression levels<sup>8</sup>. Expression levels inferred for proto-sex chromosomal precursors of Y/X gametologues using this procedure are higher than those of other proto-sex chromosome genes in all amniote lineages (Extended Data Fig. 8a), conforming to the above notion that current Y genes derive from highly expressed genes with universal (dosage-sensitive) functions. Consistently, X/Z gametologues, which are expected to generally preserve ancestral functions<sup>8</sup>, show higher expression levels than other X/Z-linked genes (Extended Data Fig. 8a).

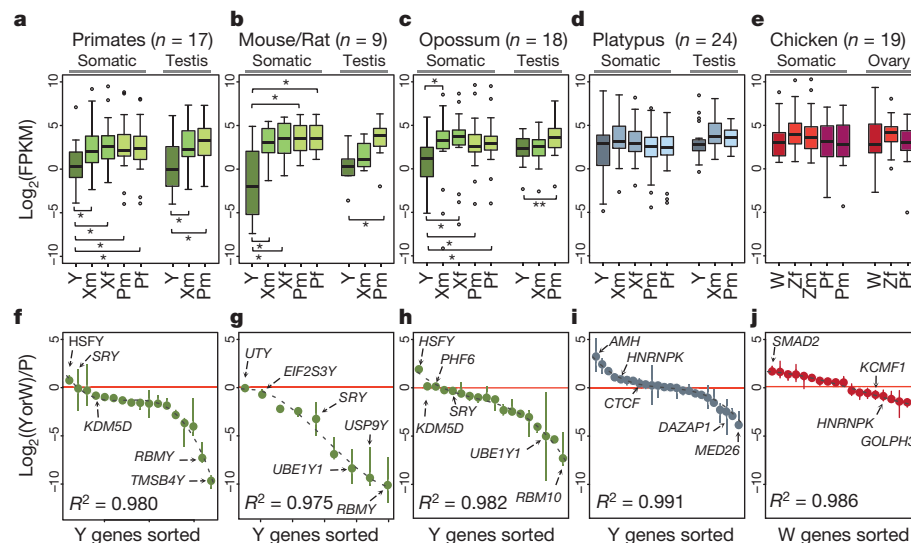
Expression levels of therian Y genes decreased since sex chromosome differentiation (corrected  $P < 0.05$ , Mann–Whitney  $U$ -test), with current (median) expression levels that are between 3.1 (opossum) to 15 (rodents) times lower than inferred ancestral expression levels of single proto-sex chromosome alleles (Fig. 2), reflecting partial regulatory decay of Y genes and/or evolution towards new functions. Further analyses show that many therian Y genes maintained ancestral patterns of usually ubiquitous expression, with interesting exceptions such as *SRY* and *AMHY*, which appear to have been sex-biased already on the proto-sex chromosomes (Fig. 3). Nevertheless, a larger number of genes evolved new expression patterns on the Y than on the X or autosomes (corrected  $P < 0.05$ , Fisher's exact test, Supplementary Table 23). All of these Y genes evolved testis-specific expression (Fig. 3), consistent with the male-limited

transmission of the Y<sup>11</sup>. Notably, therian Y genes evolved testis specificity by experiencing substantially stronger expression reduction in somatic tissues than in testis ( $P < 0.05$ , Mann–Whitney  $U$ -test; Extended Data Fig. 8b). Thus, testis specificity evolved through differential regulatory decay rather than upregulation in testis, as observed in fruitflies<sup>39</sup>. Contrary to therians, platypus Y and chicken W genes show no overall expression reduction relative to their proto-sex precursors (Fig. 2), and most of them also preserved ancestral spatial expression patterns (Fig. 3). Thus, Y/W expression preservation is particularly pronounced in monotremes and birds.

To explore cellular functions of therian testis-specific Y genes, we used the mouse as a model, assessing expression patterns of mouse Y genes across all major testicular cell types using dedicated data<sup>41</sup>. Contrary to ubiquitously expressed Y genes, testis-specific Y transcripts are restricted to three germ-cell types: mitotic germ cells (spermatogonia), meiotic cells (spermatocytes) and, especially, post-meiotic cells (round spermatids) (Extended Data Fig. 9a). Although the high expression levels of these genes in spermatids may partly reflect 'promiscuous' transcription, facilitated by overall open chromatin<sup>41</sup>, it may also indicate post-meiotic functions, consistent with the finding that the spermatid-specific *SLY* gene (Extended Data Fig. 9a) is a sex chromosome regulator in spermatids<sup>42</sup>. Detected noncoding and pseudogenetic Y transcripts are often testis specific across species and sometimes have many genomic copies (Supplementary Tables 5–17). Notably, nearly all noncoding mouse Y transcripts are specifically expressed in spermatids/spermatocytes (Extended Data Fig. 9b), probably mainly due to the permissive chromatin environment<sup>41</sup>. The lower abundance of transcripts in spermatocytes relative to spermatids is presumably due to meiotic sex chromosome inactivation, which counteracts transcriptional promiscuity<sup>41</sup>.

## Overall Y evolution and selective pressures

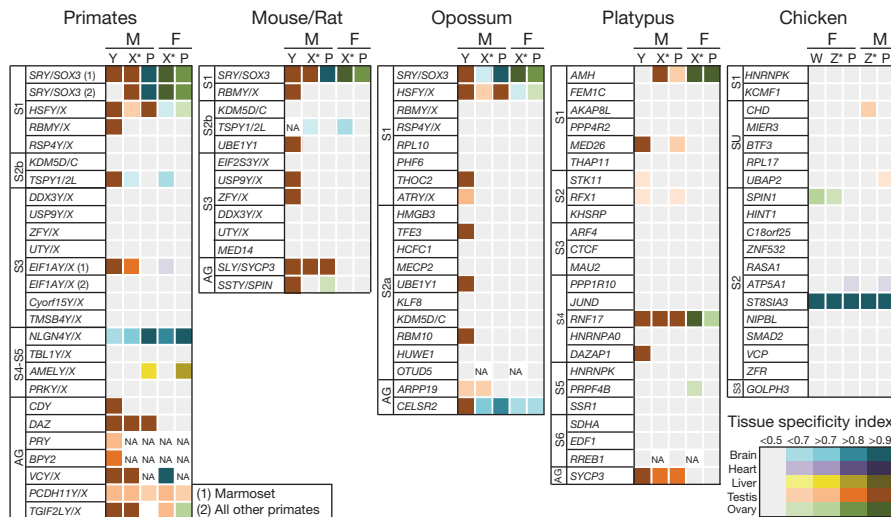
To further characterize selective signatures of Y evolution, we calculated branch-specific rates of nonsynonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitutions, which revealed substantially increased  $d_N/d_S$  values for Y/W genes compared to both X/Z gametologues and autosomal precursors, consistent with previous work<sup>43,44</sup> (Extended Data Fig. 10a). Y branch  $d_N/d_S$  is always statistically significantly below 1 and we find no evidence for positive selection, neither for the most basal Y gene branches



**Figure 2 | Expression level evolution on amniote sex chromosomes.**

**a–d**, Expression level distributions (based on medians across somatic tissues or testis) of Y genes (Y); X genes in males (Xm) and females (Xf); precursors of X/Y genes on proto-sex chromosomes in males (Pm) and females (Pf). **e**, Similar distributions for (proto-)sex chromosomes in chicken. Note: for proto-sex chromosome plots, inferred expression output values were calculated per single gene copy/allele, to assess conservation of ancestral expression levels in current single Y (W) chromosomes. Significant differences (Mann–Whitney  $U$ -test):

Benjamini–Hochberg-corrected  $*P < 0.05$ ,  $**P < 0.01$ . Error bars, maximum and minimum values, excluding outliers. **f–j**, Median current/ancestral expression level ratios for individual genes. Error bars, 95% confidence intervals. Ratios are plotted on a log<sub>2</sub> scale, that is, a ratio of 0 (non-log<sub>2</sub> ratio of 1, red line) indicates that current and ancestral expression levels are similar.  $R^2$  statistics represents the best fit to a third-order exponential curve. Gene numbers ( $n$ ) underlying the data are indicated.



**Figure 3 | Spatial expression pattern evolution on amniote sex chromosomes.** Expression patterns of Y/X genes and proto-sex chromosome (P) precursors (inferred from 1:1 autosomal orthologues in outgroups), or 'added genes' (AG) and duplicate/precursor genes (asterisk), in males and females, as assessed by the tissue-specificity index. Grey indicates ubiquitous expression; darker tones of colours indicate increasing specificity of expression in a given tissue. For some genes, homologues could not be analysed (for example, owing to lack of X copy or missing data; not applicable, NA).

nor when combining descendant branches following gametologue differentiation (Methods). Thus, although some Y genes may have been shaped by positive selection in certain species<sup>12,45</sup>, Y chromosome evolution is generally characterized by relaxed purifying selection, consistent with the impaired selection on the MSY<sup>7,43,46</sup>. Thus, Y genes probably generally preserved ancestral protein functions, consistent with the observation that even SRY may be functionally replaced by its X counterpart and functional precursor SOX3 (ref. 47) and that most X genes likely underlying Turner syndrome (impaired development due to X chromosome monosomy) not only escape X inactivation and are under strong purifying selection but also have conserved Y gametologues<sup>48</sup>. Together with our expression and simulation results, these considerations suggest that most Y genes were, at least initially, preserved because of (regulatory) dosage constraints. Some Y genes then evolved new functions in spermatogenesis or development, primarily through temporal and/or spatial expression shifts, as exemplified by the therian sex-determining gene, SRY.

## METHODS SUMMARY

We collected Illumina RNA-seq data (polyadenylated RNA fraction) for 166 tissue samples from nine mammals and two birds, and high-throughput Illumina paired-end genomic sequencing data for males and females from nine mammals and chicken. We developed a subtraction approach based on male/female RNA-seq data, Illumina genomic data and available genomes to identify and assemble Y (W) transcripts. The approach was validated using large-scale PCR and Sanger sequencing. We used genomic read coverage to predict multi-copy Y (W) genes, and male/female read coverage differences to identify X-linked genes in platypus. Phylogenetic tree reconstruction of sex-chromosome-linked genes and 1:1 autosomal orthologues from outgroup species, and associated synonymous substitutions rate analyses, were used to estimate ages of sex chromosome stratification events. We performed Monte Carlo simulations to assess non-random overlaps of Y (W) gene functions across different sex chromosome systems. RNA-seq read mapping and expression level estimation were performed using standard procedures and reference genomes to which newly identified Y (W) genes were added. To infer ancestral expression levels of Y (W) and X (Z) genes, we used 1:1 autosomal orthologues from outgroup species with different sex chromosome systems. All newly established Y and W sequences are provided in Supplementary Data 1.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2013; accepted 17 February 2014.

- Kashimada, K. & Koopman, P. Sry: the master switch in mammalian sex determination. *Development* **137**, 3921–3930 (2010).
- Wilson, M. A. & Makova, K. D. Genomic analyses of sex chromosome evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 333–354 (2009).
- Potrzebowski, L. et al. Chromosomal gene movements reflect the recent origin and biology of therian sex chromosomes. *PLoS Biol.* **6**, e80 (2008).

- Veyrunes, F. et al. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* **18**, 965–973 (2008).
- Grützner, F. et al. In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432**, 913–917 (2004).
- Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Phil. Trans. R. Soc. Lond. B* **355**, 1563–1572 (2000).
- Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Rev. Genet.* **14**, 113–124 (2013).
- Julien, P. et al. Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. *PLoS Biol.* **10**, e1001328 (2012).
- Hughes, J. F. et al. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
- Hughes, J. F. et al. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
- Skaletsky, H. et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
- Li, G. et al. Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res.* **23**, 1486–1495 (2013).
- Marais, G. A., Campos, P. R. & Gordo, I. Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol. Evol.* **2**, 347–357 (2010).
- Murphy, W. J., Pevzner, P. A. & O'Brien, S. J. Mammalian phylogenomics comes of age. *Trends Genet.* **20**, 631–639 (2004).
- Ayers, K. L. et al. RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* **14**, R26 (2013).
- Moghaddam, H. K., Pointer, M. A., Wright, A. E., Berlin, S. & Mank, J. E. W chromosome expression responds to female-specific selection. *Proc. Natl Acad. Sci. USA* **109**, 8207–8211 (2012).
- Ross, M. T. et al. The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
- Murtagh, V. J. et al. Evolutionary history of novel genes on the tammar wallaby Y chromosome: implications for sex chromosome evolution. *Genome Res.* **22**, 498–507 (2012).
- Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
- Wilson Sayres, M. A. & Makova, K. D. Genome analyses substantiate male mutation bias in many species. *Bioessays* **33**, 938–945 (2011).
- Pearks Wilkerson, A. J. et al. Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. *Genomics* **92**, 329–338 (2008).
- Chang, T. C., Yang, Y., Retzel, E. F. & Liu, W. S. Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc. Natl Acad. Sci. USA* **110**, 12373–12378 (2013).
- Weller, P. A., Critcher, R., Goodfellow, P. N., German, J. & Ellis, N. A. The human Y chromosome homologue of XG: transcription of a naturally truncated gene. *Hum. Mol. Genet.* **4**, 859–868 (1995).
- Rodríguez-Delgado, C. L., Waters, P. D., Gilbert, C., Robinson, T. J. & Graves, J. A. Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Res.* **17**, 917–926 (2009).
- Rens, W. et al. The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. *Genome Biol.* **8**, R243 (2007).
- Warren, W. C. et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
- Cutting, A., Chue, J. & Smith, C. A. Just how conserved is vertebrate sex determination? *Dev. Dyn.* **242**, 380–387 (2013).



28. Western, P. S., Harry, J. L., Graves, J. A. & Sinclair, A. H. Temperature-dependent sex determination in the American alligator: AMH precedes SOX9 expression. *Dev. Dyn.* **216**, 411–419 (1999).
29. Hattori, R. S. *et al.* A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proc. Natl Acad. Sci. USA* **109**, 2955–2959 (2012).
30. Tsend-Ayush, E. *et al.* Identification of mediator complex 26 (*Crs7*) gametologs on platypus X1 and Y5 sex chromosomes: a candidate testis-determining gene in monotremes? *Chromosome Res.* **20**, 127–138 (2012).
31. Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* **8**, e1000475 (2010).
32. Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757–762 (2010).
33. Adolfsson, S. & Ellegren, H. Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Mol. Biol. Evol.* **30**, 806–810 (2013).
34. Nam, K. & Ellegren, H. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* **180**, 1131–1136 (2008).
35. Matunis, M. J., Michael, W. M. & Dreyfuss, G. Characterization and primary structure of the poly(C)-binding heterogeneous nuclear ribonucleoprotein complex K protein. *Mol. Cell. Biol.* **12**, 164–171 (1992).
36. Wu, Z. *et al.* Targeted ubiquitination and degradation of G-protein-coupled receptor kinase 5 by the DDB1–CUL4 ubiquitin ligase complex. *PLoS ONE* **7**, e43997 (2012).
37. Smith, C. A. *et al.* The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009).
38. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29 (2000).
39. Zhou, Q. & Bachrog, D. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* **337**, 341–345 (2012).
40. Kondrashov, F. A. & Koonin, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* **20**, 287–290 (2004).
41. Soumillon, M. *et al.* Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* **3**, 2179–2190 (2013).
42. Cocquet, J. *et al.* The multicopy gene *Sly* represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol.* **7**, e1000244 (2009).
43. Wilson, M. A. & Makova, K. D. Evolution and survival on eutherian sex chromosomes. *PLoS Genet.* **5**, e1000568 (2009).
44. Berlin, S. & Ellegren, H. Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *J. Mol. Evol.* **62**, 66–72 (2006).
45. Hughes, J. F., Skaletsky, H. & Page, D. C. Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the *DAZ* (*Deleted in AZoospermia*) genes. *Bioessays* **34**, 1035–1044 (2012).
46. Charlesworth, B. Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl Acad. Sci. USA* **75**, 5618–5622 (1978).
47. Sutton, E. *et al.* Identification of *SOX3* as an XX male sex reversal gene in mice and humans. *J. Clin. Invest.* **121**, 328–341 (2011).
48. Park, C., Carrel, L. & Makova, K. D. Strong purifying selection at genes escaping X chromosome inactivation. *Mol. Biol. Evol.* **27**, 2446–2450 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank K. Harshman and the Lausanne Genomics Technology Facility for high-throughput sequencing support; I. Xenarios and the Vital-IT computational facility for computational support; S. Pääbo for great ape DNA samples; C. Roos for marmoset DNA samples; P. Jensen for chicken samples; E. Tsend-Ayush for help in determining the complete sequence of the *AMHY* gene in platypus; P. Gonzalez-Rubio for help with figure designs; M. Cardoso-Moreira, F. Carelli, A. Necsulea and M. Warnefors for comments on the manuscript; the Kaessmann group in general for discussions; and the Marmoset Genome Sequencing Consortium for making the marmoset genome assembly and annotation data available and for granting permission to use them for the analyses described in this study before publication. D.T.F. was supported by the Mexican National Council for Science and Technology (CONACyT). P.D.W. was supported by an ARC fellowship. F.G. was supported by an ARC fellowship. This research was supported by grants from the European Research Council (Starting Independent Grant: 242597, SexGenTransEvolution) and the Swiss National Science Foundation (Grant: 130287) to H.K.

**Author Contributions** D.C. performed most data processing and biological analyses. R.M. processed platypus genomic data and assessed X identity of contigs and the most likely chromosomal location of X gametologues in this species. D.T.-F. performed FISH experiments in platypus. L.F. and A.L. prepared samples and generated RNA-seq and genomic sequencing libraries. L.F. performed the large-scale PCR/Sanger sequencing validation experiments. P.D.W. provided elephant and tammar wallaby fibroblast samples and advised on these species' sex chromosome biology. F.G. provided platypus and echidna samples, supervised FISH experiments, and provided advice on the sex chromosome biology of these species. The project was supervised and originally designed by H.K. The paper was written by D.C. and H.K. with input from all authors.

**Author Information** RNA and DNA sequencing data as well as reconstructed Y/W sequences have been deposited in the Gene Expression Omnibus (GEO), Sequence Read Archive (SRA) and Transcriptome Shotgun Assembly (TSA) Database under the accession codes GSE50747, SRP029216, SRP026469 and PRJNA236159. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C. ([diegoclaudio.cortezquezada@unil.ch](mailto:diegoclaudio.cortezquezada@unil.ch)) or H.K. ([henrik.kaessmann@unil.ch](mailto:henrik.kaessmann@unil.ch)).

# Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators

Daniel W. Bellott<sup>1</sup>, Jennifer F. Hughes<sup>1</sup>, Helen Skaletsky<sup>1</sup>, Laura G. Brown<sup>1</sup>, Tatyana Pyntikova<sup>1</sup>, Ting-Jan Cho<sup>1</sup>, Natalia Koutseva<sup>1</sup>, Sara Zaghlul<sup>1</sup>, Tina Graves<sup>2</sup>, Susie Rock<sup>2</sup>, Colin Kremitzki<sup>2</sup>, Robert S. Fulton<sup>2</sup>, Shannon Dugan<sup>3</sup>, Yan Ding<sup>3</sup>, Donna Morton<sup>3</sup>, Ziad Khan<sup>3</sup>, Lora Lewis<sup>3</sup>, Christian Buhay<sup>3</sup>, Qiaoyan Wang<sup>3</sup>, Jennifer Watt<sup>3</sup>, Michael Holder<sup>3</sup>, Sandy Lee<sup>3</sup>, Lynne Nazareth<sup>3</sup>, Steve Rozen<sup>1</sup>, Donna M. Muzny<sup>3</sup>, Wesley C. Warren<sup>2</sup>, Richard A. Gibbs<sup>3</sup>, Richard K. Wilson<sup>2</sup> & David C. Page<sup>1</sup>

**The human X and Y chromosomes evolved from an ordinary pair of autosomes, but millions of years ago genetic decay ravaged the Y chromosome, and only three per cent of its ancestral genes survived. We reconstructed the evolution of the Y chromosome across eight mammals to identify biases in gene content and the selective pressures that preserved the surviving ancestral genes. Our findings indicate that survival was nonrandom, and in two cases, convergent across placental and marsupial mammals. We conclude that the gene content of the Y chromosome became specialized through selection to maintain the ancestral dosage of homologous X–Y gene pairs that function as broadly expressed regulators of transcription, translation and protein stability. We propose that beyond its roles in testis determination and spermatogenesis, the Y chromosome is essential for male viability, and has unappreciated roles in Turner's syndrome and in phenotypic differences between the sexes in health and disease.**

The human X and Y chromosomes evolved from autosomes over the past 300 million years<sup>1</sup>. Only 3% of ancestral genes survive on the human Y chromosome<sup>2,3</sup>, compared to 98% on the X chromosome<sup>4</sup>. Y-chromosome decay was initially rapid but has virtually halted over the last 25 million years, leaving a stable set of ancestral genes<sup>5–7</sup>. Mathematical models of Y-chromosome decay assume all ancestral genes are equally likely to survive. However, our initial studies of the human Y chromosome suggested that its gene content is functionally coherent<sup>8</sup>, leading us to ask whether mammalian Y chromosomes preferentially retained a subset of ancestral genes, and, if so, what qualities these surviving genes share.

Our earlier analyses<sup>8</sup> of the human Y chromosome were hampered by limited knowledge of the gene content of the ancestral autosomes. Our recent cross-species comparisons enabled us to reconstruct their gene content and identify acquired genes on the X and Y chromosomes. The human X chromosome acquired and amplified testis-expressed gene families<sup>2,4</sup>. Similarly, our comparisons of the human, chimpanzee and rhesus Y chromosomes indicated recent acquisition and amplification of testis-specific genes<sup>3,5,6</sup>. Thus, both the human X and Y chromosomes gained a specialization for male reproduction by acquiring genes that were not present on the ancestral autosomes<sup>2–4</sup>.

We excluded acquired genes to independently examine ancestral Y-linked genes for characteristics that distinguished surviving genes from genes lost to decay. Because the human, chimpanzee and rhesus Y chromosomes share nearly identical ancestral gene content, we analysed five additional mammals to enhance our ability to detect biases in the decay and survival of ancestral genes. We produced finished sequence of the ancestral portions of the Y chromosomes of marmoset (*Callithrix jacchus*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), bull (*Bos taurus*) and opossum (*Monodelphis domestica*) and compared them to the published sequences of the human, chimpanzee (*Pan troglodytes*) and rhesus macaque (*Macaca mulatta*) Y chromosomes, all eight corresponding X chromosomes and the orthologous chicken (*Gallus gallus*) autosomes as an outgroup to mammalian X and Y chromosomes. Using this expanded tree of species, we reconstructed the evolution of mammalian

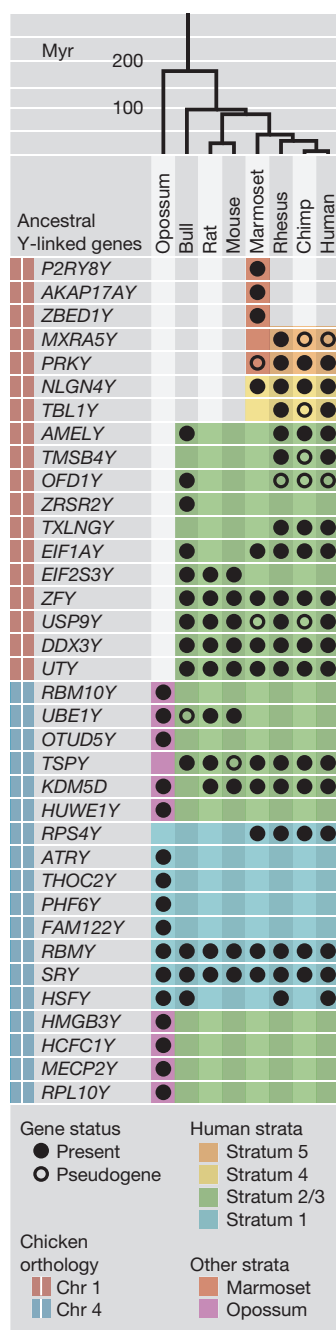
Y chromosomes from their origin to the present. We concluded that surviving Y-linked genes form a functionally coherent group enriched for dosage-sensitive, broadly expressed regulators of transcription, translation and protein stability.

We produced finished sequence using the SHIMS (single-haplotype iterative mapping and sequencing) strategy we employed on primate Y, human X and chicken Z chromosomes (Methods)<sup>2–7</sup>. These sequences comprise 17 megabases (Mb) and are accurate to about 1 nucleotide per 0.3 Mb (Supplementary Table 1, Extended Data Fig. 1 and Methods). To identify ancestral X–Y gene pairs, we searched for Y-homologues of protein-coding genes we had identified as ancestral (Supplementary Tables 2 and 3)<sup>2,5</sup>. We validated each putative gene by verifying transcriptional activity (Extended Data Fig. 2) and comparing its open reading frame to its chicken orthologue (Supplementary Data 1 and 2). We identified 36 different ancestral X–Y gene pairs across all eight species, adding 18 ancestral X–Y gene pairs to the 18 known to be present on the human, chimpanzee and rhesus Y chromosomes (Fig. 1).

## Regulatory functions of X–Y gene pairs

Seventeen years ago, we characterized human X–Y gene pairs as specialized in cellular housekeeping functions<sup>8</sup>. Since then, annotation of the human genome has increased in detail and completeness. We therefore revisited the question of functional coherence and found evidence that X–Y pair genes perform an array of regulatory functions (Fig. 2). Based on annotations of their X homologues, ancestral Y-linked genes appear to regulate each stage of the central dogma: histone lysine demethylases *KDM5D* (H3K4) and *UTY* (H3K27); the transcription factor *ZFY*, regulating stem-cell self-renewal; spliceosomal component *RBMY*; translation initiation factors *DDX3Y* and *EIF1AY*; and the deubiquitinase *USP9Y* (Fig. 2). Compared to other ancestral genes that survive on the X chromosome, X–Y pair genes are enriched for annotations such as nucleic-acid binding, transcription and translation (Extended Data Table 1, Methods and Supplementary Table 4), suggesting that X–Y pair genes can govern expression of targets throughout the genome.

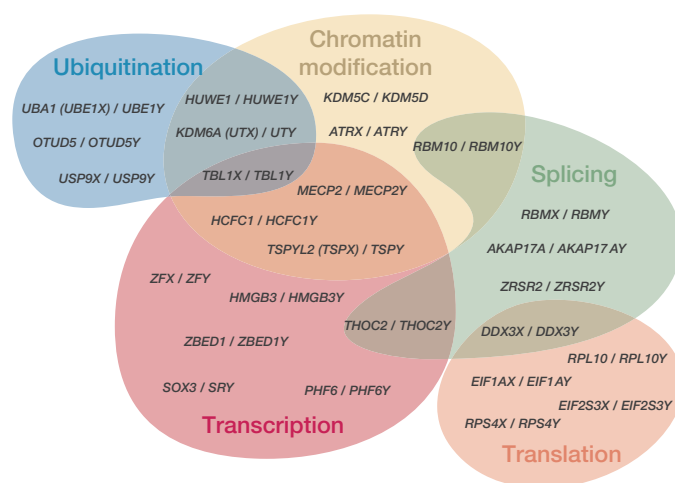
<sup>1</sup>Whitehead Institute, Howard Hughes Medical Institute, & Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. <sup>2</sup>The Genome Institute, Washington University School of Medicine, St. Louis, Missouri 63108, USA. <sup>3</sup>Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA.



**Figure 1 | Ancestral Y-linked genes by species and human X homologue location.** Ancestral Y-linked genes (filled circles) and pseudogenes (open circles) listed by the position of their X-linked homologue on the human X chromosome. The placental-specific added region (red bar) and the conserved region shared with marsupials (blue bar) of the sex chromosomes are indicated on the left. Human sex chromosome evolution was punctuated by formation of at least 4 evolutionary strata (light blue, green, yellow and orange); other strata formed independently in opossum (purple) and marmoset (red). Myr, million years.

### Convergent survival of X–Y gene pairs

To gain insight into the decay and survival of ancestral genes, we reconstructed Y chromosome evolution, taking advantage of our earlier discovery that synonymous nucleotide divergence between the X and Y sequences of each gene pair increases in stepwise fashion along the human X chromosome<sup>1,3,9</sup>. This suggested a series of discrete events, most likely inversions on the Y chromosome, that suppressed X–Y crossing over in a single region, or ‘stratum,’ without disturbing gene order on the X chromosome<sup>1,9</sup>. We used the 36 X–Y gene pairs to recalibrate previous



**Figure 2 | Regulatory annotations of X–Y pair genes.** Venn diagram depicting regulatory functions predicted for selected X–Y pair genes on basis of UniProt annotations of human X-homologue. Common alternatives to official gene symbols in parentheses.

reconstructions of evolutionary strata (Extended Data Table 2, Extended Data Figs 3–5, Methods and Supplementary Tables 2 and 5). In broad agreement with previous reconstructions<sup>1–3,9,10</sup>, we concluded that the human X and Y chromosomes evolved from ordinary autosomes through chromosomal fusion and formation of at least four strata (Fig. 3 and Methods).

Our results indicate that the stratum containing *UBE1Y* and *KDM5D* formed independently in the placental and marsupial lineages (Extended Data Fig. 4). The same set of ancestral genes became subject to genetic decay in each lineage, forming replicates of the same natural experiment. Out of the 184 ancestral genes shared between these strata, nine survived on the Y chromosome in marsupials, and three survived in placental mammals, but both lineages retained *UBE1Y* and *KDM5D* (Fig. 1, Supplementary Table 2 and Methods). The convergent survival of two ancestral genes is unlikely to occur under a model where genes survive genetic decay at random (one-tailed Fisher’s exact test,  $P < 6.25 \times 10^{-3}$ ).

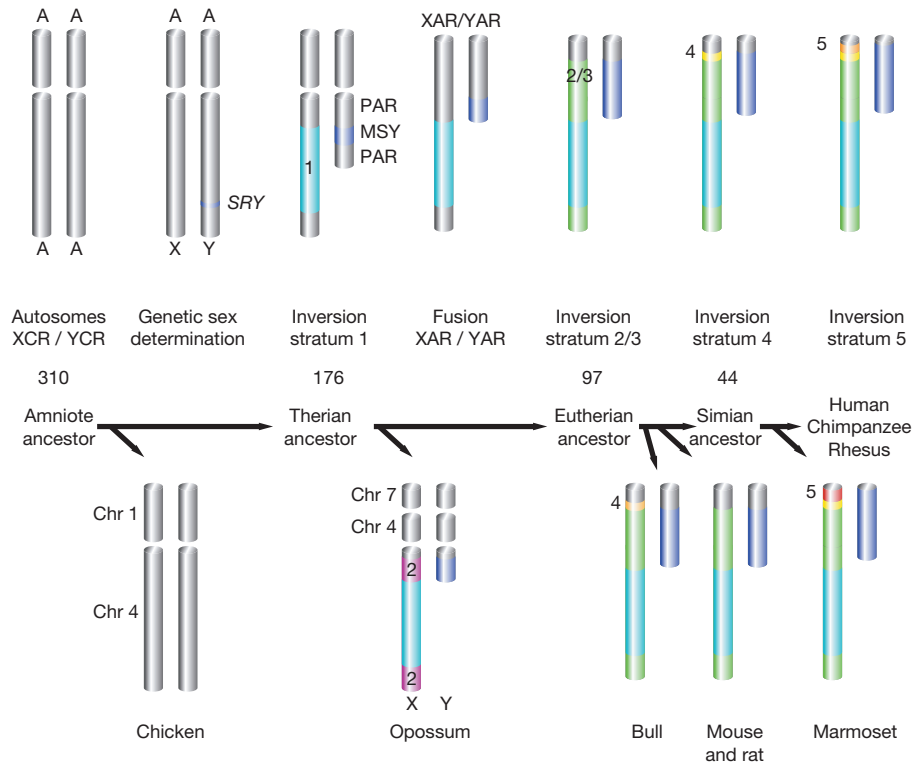
### Remarkable longevity of X–Y gene pairs

Using these recalibrated evolutionary strata, we re-examined the kinetics of genetic decay among ancestral Y-chromosome genes. Analysis of primate Y chromosomes had led us to conclude that, within a stratum, rapid gene loss was followed by stabilization at a baseline set of genes<sup>5</sup>. With five more divergent mammals, we doubled the constraints on the kinetics of gene loss during human Y chromosome evolution (Fig. 4 and Methods) and traced the stability of human Y-chromosome genes to the origin of mammals (Fig. 4). We infer that 97 million years ago, the Y chromosome of the common ancestor of placental mammals carried 18 ancestral genes from stratum 1 and stratum 2/3 (Fig. 1). Of those 18 genes, 14 survive in the human lineage (Fig. 1), and none have been lost in the last 44 million years (Fig. 4). We also examined whether ancestral Y-linked genes were stable in marsupials. Recent analyses of the tammar wallaby (*Macropus eugenii*) Y chromosome identified ten genes shared with the Tasmanian devil (*Sarcophilus harrisii*)<sup>11</sup>; we observe that all are ancestral and survive in the opossum. This suggests the opossum lineage maintained these genes over the last 78 million years<sup>12</sup>. We conclude that in both placental and marsupial lineages, some ancestral X–Y gene pairs were remarkably long lived despite rapid decay of surrounding genes.

### Two strategies preserved Y-linked genes

In light of the regulatory annotations of X–Y gene pairs, convergent survival of X–Y gene pairs in the placental and marsupial lineages, and the longevity of ancestral X–Y gene pairs across mammals, we sought





**Figure 3 | Reconstruction of human sex chromosome evolution.** Major events in the evolution of the human sex chromosomes are labelled with approximate dates in Myr. After *SRY* evolved, at least 4 evolutionary strata (light blue, green, yellow and orange) formed in the lineage leading to the human Y chromosome. Each stratum expanded the MSY (male-specific

region of the Y, deep blue) at the expense of the PAR (pseudoautosomal region, grey). Genetic decay eliminated most genes from MSY. A chromosomal fusion extended the PAR, generating conserved (XCR/YCR) and added (XAR/YAR) regions.

the evolutionary pressures that drove their survival. We had previously speculated that biases in the gene content of the human Y chromosome could arise through two evolutionary strategies: retention and amplification of testis-specific gene families, and conservation of ancestral X–Y gene pairs to maintain comparable expression between males and females<sup>8</sup>. Using the set of 639 ancestral genes reconstructed through cross-species comparisons of the human X chromosome and orthologous chicken autosomes<sup>2,4,5</sup>, we tested whether these hypotheses account for the 36 ancestral X–Y pair genes found on eight present-day Y chromosomes.

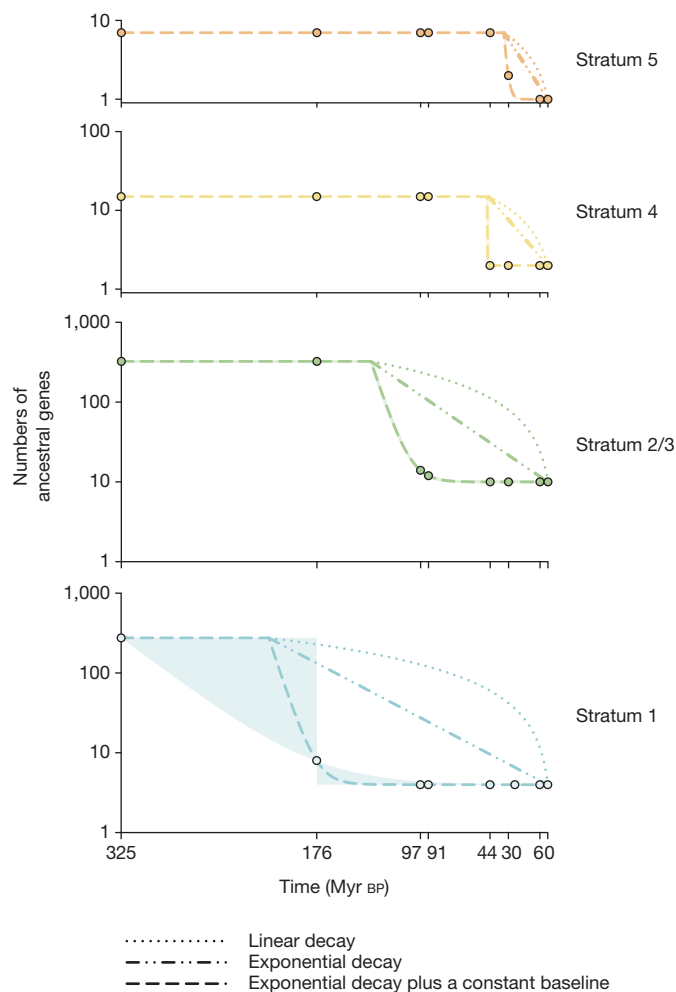
The Y chromosome was predicted to accumulate genes that enhance male reproductive fitness<sup>13</sup>, which depends upon sperm production in the adult testis. In each species we studied, ancestral genes that are amplified into multi-copy families are expressed exclusively or predominantly in the testis (Extended Data Fig. 2). However, many such genes have broadly expressed single-copy homologues on orthologous chicken autosomes, on mammalian X chromosomes, and in cases like *DDX3Y*, *EIF1AY*, *UBE1Y* and *ZFY*, on other Y chromosomes (Extended Data Fig. 2 and Supplementary Table 2). This suggests that adoption of testis-specific function preceded gene amplification.

In light of evidence that intrachromosomal gene conversion preserved testis-specific gene families in primate Y-chromosome palindromes<sup>14</sup>, we speculated that gene amplification contributed to longevity. We ranked surviving Y-linked genes by total branch length across our tree of eight species (Fig. 5a)<sup>12</sup>. Genes that are amplified in at least one species have a significantly greater branch length than those that are single copy in every species (one-tailed Mann–Whitney *U*-test,  $P < 4.27 \times 10^{-5}$ ) (Fig. 5a). This correlation remains robust when the opossum lineage, with a large number of unique single-copy genes, is excluded (one-tailed Mann–Whitney *U*-test,  $P < 5.54 \times 10^{-4}$ ). Gene families in tandem arrays show high intraspecies identity and interspecies divergence, a sign that gene conversion is more frequent than mutation in these structures (Extended Data Fig. 6). Two pairs of Y-linked genes, *RPS4Y1* and *RPS4Y2* in primates and *Zfy1* and *Zfy2* in mouse, are exceptions.

Both are physically dispersed and show no sign of recent Y–Y gene conversion (Extended Data Fig. 6). We conclude that genes specialized for male reproduction avoided genetic decay through intrachromosomal gene conversion among members of a Y-linked, multi-copy gene family.

Next, we examined whether single-copy genes on the Y chromosome survived owing to selection to preserve the correct dosage of broadly expressed genes critical to both sexes<sup>3,8,15</sup>. Most genes on the Y chromosome were lost to genetic decay, and the X chromosome evolved mechanisms to compensate for the lost dosage of Y-linked genes in males<sup>8,16,17</sup>. The Y chromosome might preferentially retain genes for which the transition state of this process, with a non-functional Y-linked gene and a functional but non-dosage-compensated X-linked homologue, was disadvantageous. Dosage-sensitive genes functioning in many tissues and cell types might be particularly sensitive to these pressures<sup>15</sup>. We re-analysed published data sets for evidence that our set of 36 X–Y pair genes systematically differ from the 603 other ancestral genes on the X chromosome with regard to dosage sensitivity<sup>18–20</sup>, breadth of expression<sup>21,22</sup> and intensity of purifying selection<sup>23</sup>.

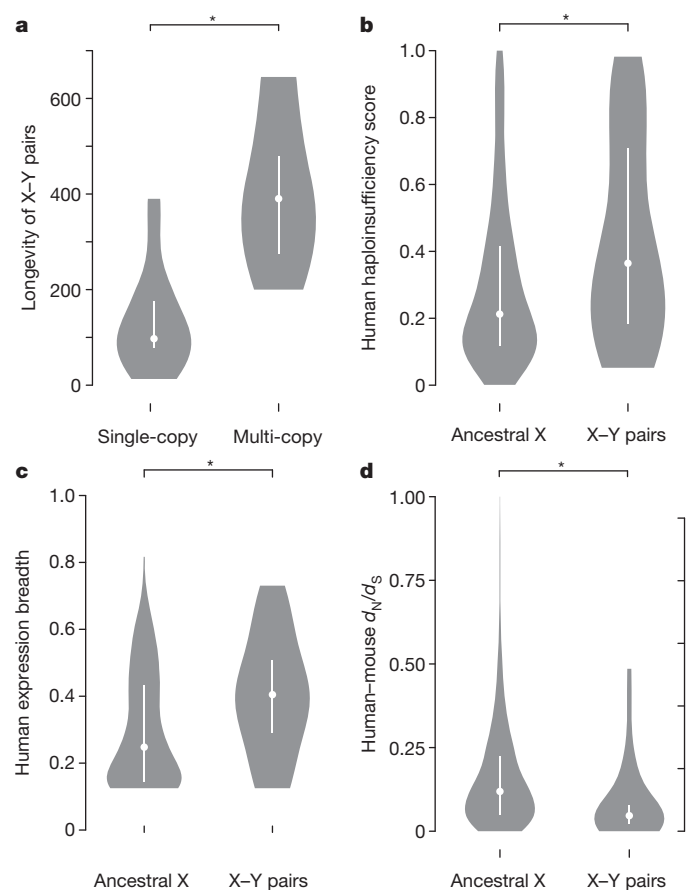
We examined whether X–Y pair genes show signs of dosage sensitivity. In humans, gene-by-gene estimates predict a greater likelihood of haploinsufficiency<sup>18</sup> for ancestral X-linked genes with surviving Y homologues compared to those lacking Y homologues (one-tailed Mann–Whitney *U*-test,  $P < 6.59 \times 10^{-3}$ ) (Fig. 5b). If surviving X–Y gene pairs maintain ancestral gene dosage, then X-linked genes with surviving Y-linked homologues should escape X inactivation. In human<sup>19</sup>, mouse<sup>20</sup>, and opossum<sup>24</sup>, data on allele-specific expression in females is informative for a subset of ancestral genes (Supplementary Table 2). In each species, a higher proportion of X-linked genes with surviving Y-linked homologues escape X-inactivation compared to those without surviving Y-linked homologues (Supplementary Table 2), and X–Y gene pairs in which the X-homologue is subject to X-inactivation have Y-homologues that show signs of functional differentiation. In humans, 12 of 14 informative



**Figure 4 | Decay of Y-linked genes to a baseline level.** Gene numbers (on a log scale on the y axis) plotted versus time (in Myr before present (Myr BP) on the x axis). Filled circles show inferred or observed gene numbers in (from left to right) Ancestral X–Y genes (before stratum formation), the MSY of common ancestor of human and opossum (176 Myr BP), bull (97 Myr BP), mouse and rat (91 Myr BP), marmoset (44 Myr BP), rhesus (30 Myr BP) and chimpanzee (6 Myr BP), and modern human MSY. Lines represent best-fit curves to data points using alternate models of decay. Exponential decay to a constant baseline provides the best fit; shaded regions represent parameters producing an equally good fit.

X–Y pair genes escape X inactivation, but only 168 of 385 remaining ancestral X genes escape (one-tailed Fisher's exact test,  $P < 1.89 \times 10^{-3}$ ) (Supplementary Table 2). The two exceptions, *TSPY* and *RBMV*, are amplified into testis-specific gene families (Extended Data Figs 2 and 6). In mouse, in which X chromosome inactivation is more complete<sup>20</sup>, four of nine informative X–Y pair genes escape X inactivation, whereas only five of 344 remaining ancestral genes escape (one-tailed Fisher's exact test,  $P < 2.36 \times 10^{-5}$ ) (Supplementary Table 2). All five exceptions (*Sry*, *Rbmy*, *Ube1y*, *Usp9y* and *Zfy*) evolved testis-specific expression in mouse (Extended Data Fig. 2). Despite differences in the mechanisms of X inactivation between placental and marsupial mammals, all eight informative opossum X–Y pair genes escape X inactivation, but only 15 of 138 remaining ancestral genes escape (one-tailed Fisher's exact test,  $P < 1.17 \times 10^{-7}$ ) (Supplementary Table 2).

The Turner's syndrome phenotype (classically associated with a 45, X karyotype, or monosomy X) suggests a strict dosage requirement for one or more sex-linked genes in humans. If dosage of X–Y pair genes is partially responsible for the Turner's syndrome phenotype, it could explain the differing features of monosomy X in humans and mice. Monosomy X in humans results in poor *in utero* viability. Less than 1 in 100 45,X



**Figure 5 | Factors in the survival of Y-linked genes.** Violin plots, white bar, interquartile range; circle, median value; asterisk, significant difference in one-tailed Mann–Whitney U-test. **a**, Multi-copy genes ( $n = 9$ ) have greater longevity than single-copy genes ( $n = 27$ ) ( $P < 4.28 \times 10^{-5}$ ). **b**, X–Y pair genes ( $n = 32$ ) have higher haploinsufficiency probability than other ancestral X genes ( $n = 478$ ) ( $P < 6.59 \times 10^{-3}$ ). **c**, X–Y pair genes ( $n = 28$ ) have broader expression across human tissues than other ancestral X genes ( $n = 383$ ) ( $P < 2.20 \times 10^{-3}$ ). **d**, X–Y pair genes ( $n = 27$ ) have lower  $d_N/d_S$  ratio than other ancestral X genes ( $n = 489$ ) ( $P < 3.39 \times 10^{-4}$ ).

conceptuses survive to term<sup>25,26</sup>. Those that do survive are often mosaic for all or part of a second sex chromosome<sup>26,27</sup>, so that variability in the Turner's syndrome phenotype may reflect variability in dosage of X–Y pair genes among tissues as well as individuals. The mouse phenotype of monosomy X is less severe; animals are small but viable and have reduced fertility<sup>28–30</sup>. This milder phenotype may reflect a dearth of genes on the mouse X chromosome that require two doses: only nine ancestral genes survive on the mouse Y chromosome (compared to 17 in human), and fewer X-linked genes escape inactivation.

Finally, human X-linked intellectual disability syndromes provide evidence for the dosage sensitivity of specific X–Y pair genes. *UTX* (also known as *KMD6A*), *KDM5C* and *NLGN4X* all have Y homologues, escape X inactivation, and appear to be haploinsufficient (Supplementary Table 2). Mutations in *UTX* cause Kabuki syndrome; both duplications and deletions result in multiple congenital anomalies and intellectual disability in males and females<sup>31</sup>. *KDM5C* is associated with X-linked intellectual disability in hemizygous males, and heterozygous females with mild intellectual disability have been reported in several families<sup>32</sup>. In both hemizygous males and heterozygous females, *NLGN4X* mutations are associated with autism spectrum disorders and learning disabilities reminiscent of the cognitive and behavioural phenotypes of Turner's syndrome<sup>33</sup>. Even the human X-homologues of X–Y gene pairs identified only in the opossum (*HCFC1*, *HUWE1* and *MECP2*) still display acute sensitivity to gene dosage. In humans each of these X-linked genes has no Y homologue and is subject to X inactivation<sup>19</sup> (Supplementary

Table 2). Nevertheless, a non-coding mutation causing overexpression of *HCFC1*, as well as duplications of *HUWE1* and *MECP2*, have been implicated in X-linked intellectual disability in human males<sup>34–36</sup>. Thus, even though the human Y-homologues of *HCFC1*, *HUWE1* and *MECP2* were lost and the surviving X-homologues have evolved dosage compensation, their gene dosage remains tightly constrained.

X–Y pair genes functioning across many tissues and cell types may face additional selective constraints that prevent both loss of the Y-linked gene and evolution of a dosage-compensated gene on the X chromosome. In all eight species, single-copy Y-linked genes are broadly expressed across adult tissues (Extended Data Fig. 2), with two major exceptions, in which both members of the X–Y pair share ancestrally restricted expression. *AMELY*, whose orthologue disappeared in the toothless avian lineage, is expressed only in developing tooth buds<sup>37</sup>; and *HSFY*, which is testis specific, and has a chicken orthologue that is predominantly expressed in testis. In chicken, the autosomal orthologues of mammalian X–Y pair genes have significantly broader expression across adult tissues than do the orthologues of ancestral genes that survive only on the X chromosome, and X–Y pair genes maintain this broader expression across mammals (one-tailed Mann–Whitney *U*-test, chicken  $P < 3.38 \times 10^{-3}$ ; human  $P < 2.20 \times 10^{-3}$ ; rhesus  $P < 1.39 \times 10^{-7}$ ; mouse  $P < 4.74 \times 10^{-8}$ ; rat  $P < 4.63 \times 10^{-6}$ ; bull  $P < 1.20 \times 10^{-5}$ ) (Fig. 5c and Supplementary Table 2). This breadth of expression extends to the earliest stages of development. Relative to other X-linked ancestral genes on the X chromosome, X–Y pair genes are enriched for genes upregulated after the onset of zygotic gene activation in a time course of human, mouse and bovine pre-implantation development (one-tailed Fisher's exact test, human  $P < 2.13 \times 10^{-2}$ ; mouse  $P < 5.93 \times 10^{-4}$ ; bull  $P < 1.37 \times 10^{-2}$ ) (Supplementary Table 2). X–Y pair genes are more broadly expressed than other ancestral genes that survive on the X chromosome, across many tissues and developmental time.

Unlike the testis-expressed multi-copy gene families, the broadly expressed, dosage-sensitive single-copy genes of the Y chromosome cannot avoid genetic decay through intrachromosomal gene conversion, and must rely on purifying selection. Our previous survey of human sequence variation among the single-copy genes on the Y chromosome showed that natural selection operated effectively to preserve the amino acid sequences of Y-linked genes in the human lineage<sup>38</sup>. If X–Y gene pairs are haploinsufficient, alleles that alter the function of the X-linked homologues should be detrimental in both males and females. We examined Ensembl human–mouse orthologue alignment data for evidence that the X-linked homologues of X–Y gene pairs were subject to strong purifying selection<sup>23</sup>. Relative to other ancestral genes on the X chromosome, the X-linked homologues of X–Y gene pairs have a reduced ratio of non-synonymous to synonymous substitution rates ( $dN/dS$ ) (one-tailed Mann–Whitney *U*-test,  $P < 3.39 \times 10^{-4}$ ) (Fig. 5d). We conclude that these broadly expressed, dosage-sensitive X–Y pair genes are under more intense purifying selection than their neighbours on the X chromosome.

### Human Y genes ensure male viability

We conclude that the longevity of many Y-linked genes is due to selection to maintain expression, in males, of dosage-sensitive, broadly expressed X–Y gene pairs at levels comparable to their autosomal ancestors. This model predicts that members of surviving single-copy X–Y gene pairs should be functionally interchangeable. Indeed, the human Y-linked genes *RPS4Y1* and *DDX3Y* are functionally interchangeable with their X homologues *in vitro*<sup>39,40</sup>, and although the histone demethylase domain of the mouse Y-linked gene *Uty* appears to be inactive, mouse *Utx* and *Uty* are functionally redundant during mouse embryonic development<sup>41–43</sup>.

Previous observations suggest that the selective pressures that maintained these Y-linked genes remain strong in the human lineage; about 99% of human 45,X conceptuses are inviable, and those that survive to term are often mosaic for all or part of a second sex chromosome<sup>25–27</sup>. Therefore, we also conclude that the broadly expressed, dosage-sensitive genes of the human Y chromosome—along with their X-homologues,

which escape X chromosome inactivation—are collectively haplolethal. We propose that, as a set, these dozen Y-linked genes are essential for the viability of 46,XY fetuses (Methods and Extended Data Fig. 7). Thus we propose a third organismal function of the human Y chromosome: that it carries single-copy genes that ensure male viability. This is distinct from the human Y chromosome's more widely appreciated roles in testis determination through *SRY* and sperm production through ampliconic gene families.

### Sex differences in health and disease

All of the myriad differences between human males and females—from anatomy to disease susceptibility—arise from differences in the genes of the X and Y chromosomes that appeared as these chromosomes diverged in gene content from their autosomal ancestors. Of the 17 surviving ancestral genes on the human Y chromosome, four (*SRY*, *RBMY*, *TSPY*, and *HSFY*) have clearly diverged in function from their X homologues (*SOX3*, *RBMX*, *TSPX* and *HSEFX*) to play male-specific roles in reproductive development or gametogenesis. Because all genes on the Y chromosome were exposed to selection only in males, even widely expressed ancestral genes may exhibit subtle functional differences from their X-linked homologues. Particularly worthy of consideration are eight global regulators of gene activity that exist as X-encoded and Y-encoded (male-specific) protein isoforms in diverse human tissues: *UTX/UTY*, *EIF1AX/EIF1AY*, *ZFX/ZFY*, *RPS4X/RPS4Y1*, *KDM5C/KDM5D*, *DDX3X/DDX3Y*, *USP9X/USP9Y* and *TBL1X/TBL1Y*. These exemplify a fundamental sexual dimorphism, at a biochemical level, throughout the human body, that derives directly from genetic differences between the X and Y chromosomes. It will surely be of interest to determine whether this dimorphism has a role in diseases, outside the reproductive tract, that occur with greater frequency or severity in males or females.

### METHODS SUMMARY

We used the SHIMS (single-haplotype iterative mapping and sequencing) strategy to assemble a path of sequenced clones for each organism (Methods). Contigs were ordered and oriented by radiation hybrid mapping using RHMAPPER 1.22 (ref. 44) and extended metaphase and interphase FISH, as previously described<sup>45</sup>.

We validated transcription of predicted genes by reverse-transcriptase polymerase chain reaction and capillary sequencing, as well as 454 sequencing of testis complementary DNA (cDNA), as previously described<sup>5</sup>.

We relied on Ensembl<sup>23</sup> version 70 to identify 1:1 orthologues between human, chimpanzee, rhesus, marmoset, mouse, rat and bull X chromosomes, the opossum X chromosomes and autosomes, and chicken autosomes, but manually reviewed cases where simple 1:1 orthologues were not clear<sup>2,4</sup>. Within each stratum, we identified X–Y pair genes as ancestral if their X-linked or autosomal orthologues were syntenic in an outgroup.

For each species, we aligned each X–Y pair and calculated  $d_N$ ,  $d_S$ , and  $d_N/d_S$  using PAML<sup>46</sup> to identify evolutionary strata. For cross-species phylogenetic analysis, we generated multiple alignments in MUSCLE<sup>47</sup> and used these alignments to generate a tree with 100 bootstrap replicates using DNAML in PHYLIP<sup>48</sup>. Within each stratum, we modelled gene loss as previously described<sup>6</sup>.

To calculate longevity, we summed all branch lengths in the most parsimonious tree from each of the species where a gene is present to the last common ancestor before stratum formation.

We mapped published functional annotation data<sup>18–23</sup> onto our set of ancestral genes. We identified pre-implantation expressed genes as previously described<sup>22</sup>. For expression breadth<sup>21</sup>, we normalized the expression of each X-linked gene to the highest reads per kilobase per million reads (RPKM) in any tissue, and took the average expression across all tissues. We used PANTHER<sup>49</sup> to calculate the enrichment of Gene Ontology terms in X–Y gene pairs relative to the ancestral X, and used UniProt annotations<sup>50</sup> to identify X–Y pair genes involved in regulatory processes.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 November 2013; accepted 6 March 2014.

1. Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).



2. Bellott, D. W. *et al.* Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612–616 (2010).
3. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
4. Mueller, J. L. *et al.* Independent specialization of the human and mouse X chromosomes for the male germline. *Nature Genet.* **45**, 1083–1087 (2013).
5. Hughes, J. F. *et al.* Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012).
6. Hughes, J. F. *et al.* Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536–539 (2010).
7. Hughes, J. F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100–103 (2005).
8. Lahn, B. T. & Page, D. C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
9. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
10. Watson, J. M., Spencer, J. A., Riggs, A. D. & Graves, J. A. The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. *Proc. Natl Acad. Sci. USA* **87**, 7125–7129 (1990).
11. Murtagh, V. J. *et al.* Evolutionary history of novel genes on the tammar wallaby Y chromosome: implications for sex chromosome evolution. *Genome Res.* **22**, 498–507 (2012).
12. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971–2972 (2006).
13. Fisher, R. A. The evolution of dominance. *Biol. Rev. Camb. Philos. Soc.* **6**, 345–368 (1931).
14. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
15. Kaiser, V. B., Zhou, Q. & Bachtrog, D. Nonrandom gene loss from the *Drosophila miranda* neo-Y chromosome. *Genome Biol. Evol.* **3**, 1329–1337 (2011).
16. Jegalian, K. & Page, D. C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776–780 (1998).
17. Ohno, S. *Sex Chromosomes and Sex-linked Genes* (Springer-Verlag, 1967).
18. Huang, N., Lee, I., Marcotte, E. M. & Hurles, M. E. Characterising and predicting haploinsufficiency in the human genome. *PLoS Genet.* **6**, e1001154 (2010).
19. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).
20. Yang, F., Babak, T., Shendure, J. & Distche, C. M. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res.* **20**, 614–622 (2010).
21. Merkin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2012).
22. Xie, D. *et al.* Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. *Genome Res.* **20**, 804–815 (2010).
23. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42**, D749–D755 (2014).
24. Wang, X., Douglas, K. C., Vandenberg, J. L., Clark, A. G. & Samollow, P. B. Chromosome-wide profiling of X-chromosome inactivation and epigenetic states in fetal brain and placenta of the opossum, *Monodelphis domestica*. *Genome Res.* **24**, 70–83 (2014).
25. Cockwell, A., MacKenzie, M., Youings, S. & Jacobs, P. A cytogenetic and molecular study of a series of 45,X fetuses and their parents. *J. Med. Genet.* **28**, 151–155 (1991).
26. Hook, E. B. & Warburton, D. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum. Genet.* **64**, 24–27 (1983).
27. Hassold, T., Benham, F. & Leppert, M. Cytogenetic and molecular analysis of sex-chromosome monosomy. *Am. J. Hum. Genet.* **42**, 534–541 (1988).
28. Burgoyne, P. S., Tam, P. P. & Evans, E. P. Retarded development of XO conceptuses during early pregnancy in the mouse. *J. Reprod. Fertil.* **68**, 387–393 (1983).
29. Burgoyne, P. S. & Baker, T. G. Oocyte depletion in XO mice and their XX sibs from 12 to 200 days post partum. *J. Reprod. Fertil.* **61**, 207–212 (1981).
30. Burgoyne, P. S., Evans, E. P. & Holland, K. XO monosomy is associated with reduced birthweight and lowered weight gain in the mouse. *J. Reprod. Fertil.* **68**, 381–385 (1983).
31. Lindgren, A. M. *et al.* Haploinsufficiency of KDM6A is associated with severe psychomotor retardation, global growth restriction, seizures and cleft palate. *Hum. Genet.* **132**, 537–552 (2013).
32. Rujirabanjerd, S. *et al.* Identification and characterization of two novel *JARID1C* mutations: suggestion of an emerging genotype-phenotype correlation. *Eur. J. Hum. Genet.* **18**, 330–335 (2010).
33. Lawson-Yuen, A., Saldivar, J. S., Sommer, S. & Picker, J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. *Eur. J. Hum. Genet.* **16**, 614–618 (2008).
34. Huang, L. *et al.* A noncoding, regulatory mutation implicates *HCFC1* in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **91**, 694–702 (2012).
35. Ramocki, M. B., Tavyev, Y. J. & Peters, S. U. The *MECP2* duplication syndrome. *Am. J. Med. Genet. A* **152A**, 1079–1088 (2010).
36. Froyen, G. *et al.* Copy-number gains of *HUWE1* due to replication- and recombination-based rearrangements. *Am. J. Hum. Genet.* **91**, 252–264 (2012).
37. Lau, E. C., Mohandas, T. K., Shapiro, L. J., Slavkin, H. C. & Snead, M. L. Human and mouse amelo-genin gene loci are on the sex chromosomes. *Genomics* **4**, 162–168 (1989).
38. Rozen, S., Marszalek, J. D., Alagappan, R. K., Skaletsky, H. & Page, D. C. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. *Am. J. Hum. Genet.* **85**, 923–928 (2009).
39. Watanabe, M., Zinn, A. R., Page, D. C. & Nishimoto, T. Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nature Genet.* **4**, 268–271 (1993).
40. Sekiguchi, T., Iida, H., Fukumura, J. & Nishimoto, T. Human DDX3Y, the Y-encoded isoform of RNA helicase DDX3, rescues a hamster temperature-sensitive ET24 mutant cell line with a DDX3X mutation. *Exp. Cell Res.* **300**, 213–222 (2004).
41. Weststead, G. G. *et al.* X-linked H3K27me3 demethylase Utx is required for embryonic development in a sex-specific manner. *Proc. Natl Acad. Sci. USA* **109**, 13004–13009 (2012).
42. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* **8**, e1002964 (2012).
43. Lee, S., Lee, J. W. & Lee, S. K. UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev. Cell* **22**, 25–37 (2012).
44. Slonim, D., Kruglyak, L., Stein, L. & Lander, E. Building human genome maps with radiation hybrids. *J. Comput. Biol.* **4**, 487–504 (1997).
45. Saxena, R. *et al.* The *DAZ* gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nature Genet.* **14**, 292–299 (1996).
46. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556 (1997).
47. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
48. Felsenstein, J. PHYLIP - phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166 (1989).
49. Thomas, P. D. *et al.* PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).
50. The UniProt Consortium Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **41**, D43–D47 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank W. J. Murphy, E. Owens and J. E. Womak for generating radiation hybrid panels and for assistance in mapping; L. Lyons and W.J.M. for providing the rhesus radiation hybrid panel; A. Kaur for a rhesus cell line; S. Austad, P. Hornsby and S. Tardif for marmoset cell lines; M. Brown for rat cell lines; J.E.W. for bull fibroblasts; W. Johnson and S. O'Neil for rhesus tissues samples; W.J., S.O. and S.T. for marmoset tissue samples; M. Turner for rat tissue samples; J.E.W. for bull tissue samples; P. Samollow for opossum tissue samples; E. Vallender for Tamarin DNA; B. Chowdhary and T. Raudsepp for FISH experiments in the bull; C. Friedman and B. Trask for flow-sorted marmoset Y chromosomes; B.T. for sizing rat Y chromosomes; C. Burge for permission to assemble transcriptome data from SRR594455, SRR594463 and SRR594508; J. Alföldi for permission to assemble transcriptome data from SRR500909; R.B. Norgren for permission to assemble transcriptome data from SRR544870; and A. Godfrey, Y. Hu and B. Lesch for comments on the manuscript. Supported by National Institutes of Health and Howard Hughes Medical Institute.

**Author Contributions** D.W.B., J.F.H., H.S., S. Rozen, W.C.W., R.A.G., R.K.W. and D.C.P. planned the project. J.F.H., H.S., L.G.B., T.-J.C., N.K. and S.Z. performed BAC mapping, radiation hybrid mapping and real-time polymerase chain reaction analyses. T.G., S. Rock, C.K., R.S.F., S.D., Y.D., D.M., Z.K., L.L., C.B., Q.W., J.W., M.H., S.L., L.N. and D.M.M. were responsible for BAC sequencing. D.W.B., J.F.H. and H.S. performed comparative sequence analyses. T.P. performed FISH analyses. D.W.B. and D.C.P. wrote the paper.

**Author Information** The cDNA sequences of Y-linked genes and their X-linked homologs have been deposited in GenBank (<http://www.ncbi.nlm.nih.gov>) under accession numbers FJ526999–FJ527008, FJ627275, FJ627276, FJ627278, FJ659845, FJ959389, GQ253467–GQ253475, GQ338825, GU304599–GU304603, GU304606, GU304607, JF487792–JF487795, JF827151, JF827152, JN086997, JN585955, JN585956, JQ313990–JQ313992 and BioProject PRJNA221163. The 454 and Illumina testis cDNA sequences have been deposited in GenBank under accession numbers SRX335333, SRX335335, SRX335470, SRX335472, SRX335475–SRX335477, SRX358238 and SRX359414. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.B. (bellott@wi.mit.edu).

# Superconducting quantum circuits at the surface code threshold for fault tolerance

R. Barends<sup>1\*</sup>, J. Kelly<sup>1\*</sup>, A. Megrant<sup>1</sup>, A. Veitia<sup>2</sup>, D. Sank<sup>1</sup>, E. Jeffrey<sup>1</sup>, T. C. White<sup>1</sup>, J. Mutus<sup>1</sup>, A. G. Fowler<sup>1,3</sup>, B. Campbell<sup>1</sup>, Y. Chen<sup>1</sup>, Z. Chen<sup>1</sup>, B. Chiaro<sup>1</sup>, A. Dunsworth<sup>1</sup>, C. Neill<sup>1</sup>, P. O'Malley<sup>1</sup>, P. Roushan<sup>1</sup>, A. Vainsencher<sup>1</sup>, J. Wenner<sup>1</sup>, A. N. Korotkov<sup>2</sup>, A. N. Cleland<sup>1</sup> & John M. Martinis<sup>1</sup>

A quantum computer can solve hard problems, such as prime factoring<sup>1,2</sup>, database searching<sup>3,4</sup> and quantum simulation<sup>5</sup>, at the cost of needing to protect fragile quantum states from error. Quantum error correction<sup>6</sup> provides this protection by distributing a logical state among many physical quantum bits (qubits) by means of quantum entanglement. Superconductivity is a useful phenomenon in this regard, because it allows the construction of large quantum circuits and is compatible with microfabrication. For superconducting qubits, the surface code approach to quantum computing<sup>7</sup> is a natural choice for error correction, because it uses only nearest-neighbour coupling and rapidly cycled entangling gates. The gate fidelity requirements are modest: the per-step fidelity threshold is only about 99 per cent. Here we demonstrate a universal set of logic gates in a superconducting multi-qubit processor, achieving an average single-qubit gate fidelity of 99.92 per cent and a two-qubit gate fidelity of up to 99.4 per cent. This places Josephson quantum computing at the fault-tolerance threshold for surface code error correction. Our quantum processor is a first step towards the surface code, using five qubits arranged in a linear array with nearest-neighbour coupling. As a further demonstration, we construct a five-qubit Greenberger–Horne–Zeilinger state<sup>8,9</sup> using the complete circuit and full set of gates. The results demonstrate that Josephson quantum computing is a high-fidelity technology, with a clear path to scaling up to large-scale, fault-tolerant quantum circuits.

Fault tolerance in the surface code is achieved by placing physical qubits in a checkerboard pattern, with white squares representing data and black squares representing measurement qubits that detect errors. To perform this detection, each measurement qubit needs to interact with its four neighbouring data qubits. All that is needed for these interactions are single- and two-qubit gates with sufficiently high fidelity. The high fidelity demonstrated here is achieved through a combination of coherent qubits, a straightforward interconnection architecture and a novel implementation of the two-qubit controlled-phase entangling gate. The controlled-phase gate uses a fast but adiabatic qubit frequency tuning that minimizes error<sup>10</sup>.

Here the tuneable nature of the qubits and their entangling gates provides both high fidelity and fast control. Previous demonstrations of two-qubit gates achieving >99% fidelity used fixed-frequency qubits: systems based on nuclear magnetic resonance and ion traps have shown two-qubit gates with fidelities of 99.5% (ref. 11) and 99.3% (ref. 12). Recently, for a five-qubit ion trap<sup>13</sup> and a three-qubit superconducting system<sup>14</sup>, two-qubit entangling gate fidelities of 95% and 96% were reported.

Superconductivity allows for the construction of large quantum integrated circuits as the electrons are condensed into a single macroscopic quantum state. We have designed a processor to test our ability to implement the surface code; it consists of five cross-shaped transmon qubits (Xmons) with nearest-neighbour coupling, arranged in a linear array (Fig. 1). The Xmon qubit<sup>15</sup> offers a nodal approach to connectivity while

maintaining a high level of coherence (see Supplementary Information for decoherence times). Here the four legs of the cross allow for a natural segmentation of the design into coupling, control and readout. We choose a modest inter-qubit capacitive coupling strength of  $g/2\pi = 30$  MHz and use alternating qubit idle frequencies of 5.5 and 4.7 GHz, enabling a controlled-phase gate in 40 ns when two qubits are brought near resonance, while minimizing the effective coupling to 0.3 MHz when the qubits are at their idle points. Rotations around the X and Y axes in the Bloch sphere representation are performed using pulses on the microwave (XY) line, whereas Z-axis rotations, which control the phase of the quantum state, are achieved by a flux-bias current on the frequency-control (Z) line. We use a dispersive measurement method<sup>16</sup> whereby each qubit is coupled to a readout resonator with a distinct resonance frequency, enabling simultaneous readout using frequency-domain multiplexing through a single coplanar waveguide<sup>17</sup>. The modularity of this architecture makes it straightforward to integrate more qubits in the circuit.

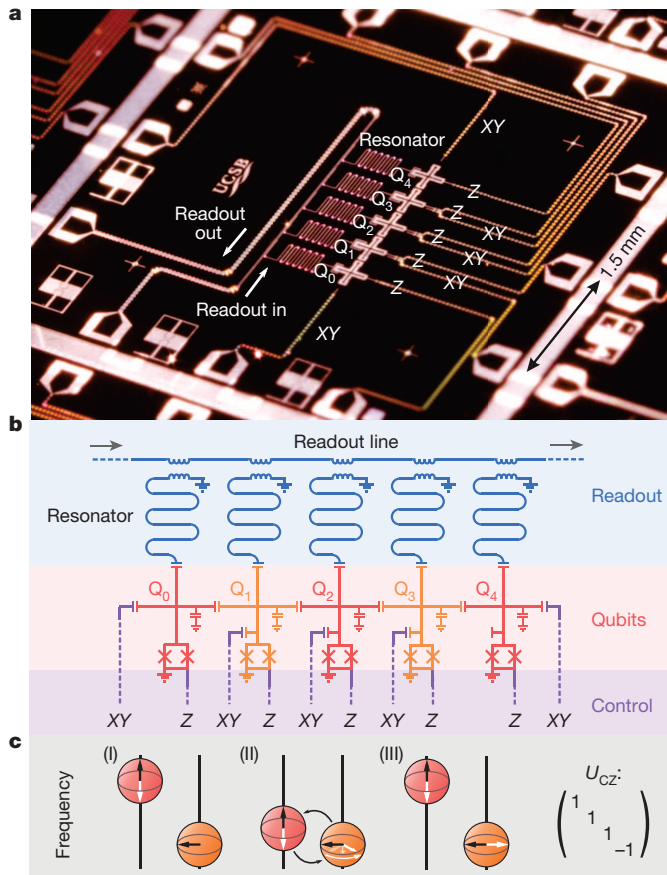
We characterize our gate fidelities using Clifford-based randomized benchmarking<sup>11,18,19</sup>. The Clifford group is a set of rotations that evenly samples the Hilbert space, and the benchmarking thus averages across errors. For the single-qubit case, the Clifford gates (which we henceforth refer to simply as Cliffords) comprise  $\pi$ ,  $\pi/2$  and  $2\pi/3$  rotations (Supplementary Information). In randomized benchmarking, a logic gate is characterized by measuring its performance when it is interleaved with many random sequences of gates, making the measured fidelity resilient to state preparation and measurement errors. We first perform a control experiment on a ground-state qubit by generating a random sequence of  $m$  Cliffords; appending the unique recovery Clifford ( $C_r$ ) that inverts the sequence; and averaging the experimental sequence fidelity, the final ground-state population, over  $k$  different sequences<sup>19,20</sup>. The resulting reference sequence fidelity,  $F_{\text{ref}}$ , is fitted to  $F_{\text{ref}} = Ap_{\text{ref}}^m + B$ , where  $p_{\text{ref}}$  is the sequence decay, and state preparation and measurement errors are captured in the parameters  $A$  and  $B$ . The average error per Clifford of the reference is given by  $r_{\text{ref}} = (1 - p_{\text{ref}})(d - 1)/d$ , with  $d = 2^{N_{\text{qubits}}}$ . We then measure the fidelity of a specific gate by interleaving it with  $m$  random Cliffords. The sequence decay,  $p_{\text{gate}}$ , gives the gate error,  $r_{\text{gate}} = (1 - p_{\text{gate}}/p_{\text{ref}})(d - 1)/d$ .

The benchmarking results for the single-qubit gates are shown in Fig. 2. We generate the Cliffords using microwave pulses, from a basis set of  $\pi$  and  $\pi/2$  rotations around the X and Y axes (Supplementary Information). We benchmark X- and Y-axis  $\pi$  and  $\pi/2$  rotations, the Hadamard gate (implemented with a  $\pi/2$  Y rotation (Y/2) followed by a  $\pi$  X rotation (X)) and Z-axis rotations. From the data in Fig. 2, we extract the individual gate fidelities listed in the key. We find an average fidelity of 99.92% over all gates and qubits (Supplementary Information). The highest fidelities are achieved by optimizing the pulse amplitude and frequency, and minimizing two-state leakage<sup>21</sup>.

We have also measured the performance when simultaneously operating nearest-neighbour or next-nearest-neighbour qubits<sup>22</sup>, with the

<sup>1</sup>Department of Physics, University of California, Santa Barbara, California 93106, USA. <sup>2</sup>Department of Electrical Engineering, University of California, Riverside, California 92521, USA. <sup>3</sup>Centre for Quantum Computation and Communication Technology, School of Physics, The University of Melbourne, Victoria 3010, Australia.

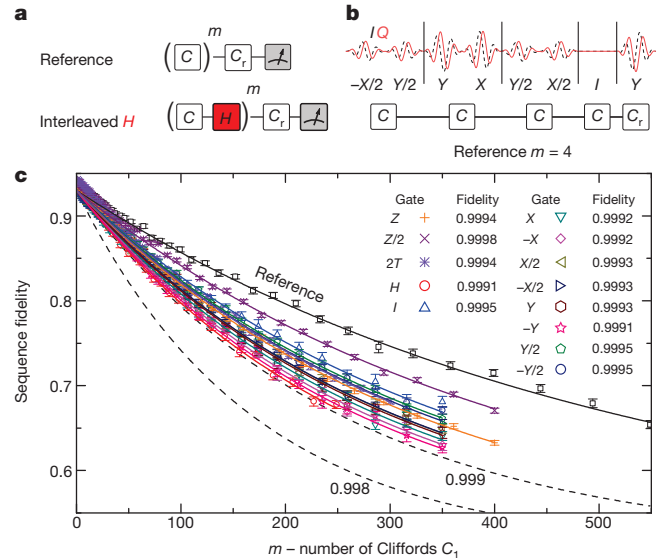
\*These authors contributed equally to this work.



**Figure 1 | Architecture.** **a**, Optical image of the integrated Josephson quantum processor, consisting of aluminium (dark) on sapphire (light). The five cross-shaped devices (Q<sub>0</sub>–Q<sub>4</sub>) are the Xmon variant of the transmon qubits<sup>30</sup>, placed in a linear array. To the left of the qubits are five meandering coplanar waveguide resonators used for individual state readout. Control wiring is brought in from the contact pads at the edge of the chip, ending at the right of the qubits. **b**, Circuit diagram. Our architecture uses direct, nearest-neighbour coupling of the qubits (red/orange), made possible by the nodal connectivity of the Xmon qubit. Using a single readout line, each qubit can be measured using frequency-domain multiplexing (blue). Individual qubits are driven through capacitively coupled microwave control lines (XY), and frequency control is achieved through inductively coupled d.c. lines (Z) (violet). **c**, Schematic representation of an entangling operation using a controlled-phase gate with unitary representation  $U_{CZ}$ : (I) qubits at rest, at distinct frequencies with minimal interaction; (II) when brought near resonance, the state-dependent frequency shift brings about a rotation conditional on the qubit states; (III) qubits are returned to their rest frequency.

qubits at dissimilar idle frequencies to minimize coupling. The fidelities are essentially unchanged, with small added errors of  $< 10^{-4}$  (Supplementary Information), showing a high degree of addressability for this architecture.

The two-qubit controlled-phase gate is implemented by tuning one qubit in frequency along a ‘fast adiabatic’ trajectory<sup>10</sup> that takes the two-qubit state  $|11\rangle$  close to the avoided level crossing with the state  $|02\rangle$  (ref. 23), yielding a state-dependent relative phase shift (Fig. 3a). This implementation is the natural choice for weakly anharmonic, frequency-tunable qubits, because the other computational states are left unchanged<sup>123–25</sup>. It is advantageous that the controlled-phase gate is adiabatic as well as fast. An adiabatic trajectory is easily optimized and allows for leakage into the non-computational state  $|02\rangle$  to be exponentially suppressed with gate duration, because slower gates are less likely to lead to undesired transitions<sup>10</sup>. Having a fast controlled-phase gate minimizes the accumulation of errors from decoherence and unwanted entanglement with other circuit elements, which is favourable for fault tolerance.



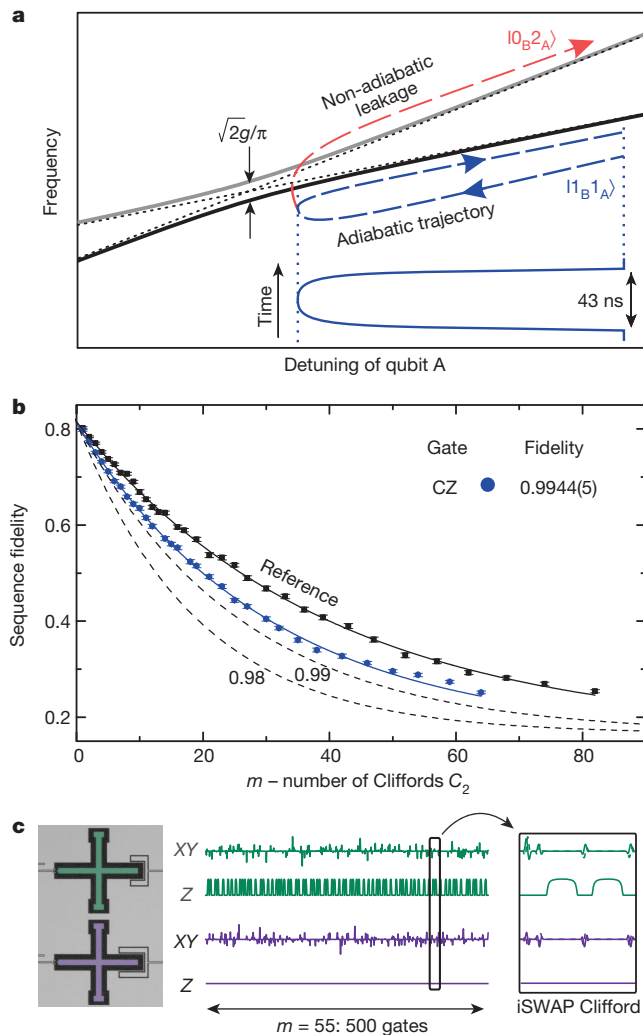
**Figure 2 | Single-qubit randomized benchmarking.** **a**, A reference experiment is performed by generating a sequence of  $m$  random Cliffords, which are inverted by the recovery Clifford,  $C_r$ . A specific gate,  $H$ , is tested using a sequence that interleaves  $H$  with  $m$  random Cliffords. The difference between interleaved and reference decay gives the gate fidelity. **b**, Representative pulse sequence for a set of four Cliffords and their recovery, generated with  $\pi$  and  $\pi/2$  rotations about  $X$  and  $Y$ , displaying both the real ( $I$ ) and imaginary ( $Q$ ) microwave pulse envelopes before up-conversion by quadrature mixing to the qubit frequency. **c**, Randomized benchmarking measurement for the set of single-qubit gates for qubit Q<sub>2</sub>, plotting reference and interleaved sequence fidelities as functions of the length,  $m$ ; the fidelity for each value of  $m$  was measured for  $k = 40$  different sequences. The fit to the reference set yields an average error per Clifford of  $r_{\text{ref}} = 0.0011$ , consistent with an average gate fidelity of  $1 - r_{\text{ref}}/1.875 = 0.9994$  (Supplementary Information). The dashed lines indicate the thresholds for exceeding gate fidelities of 0.998 and 0.999. The fidelities for the single-qubit gates are tabulated in the key. We find that all gates have a fidelity greater than 0.999. The error bars on the data points are the standard deviations from the mean. The uncertainty in gate fidelity is typically  $5 \times 10^{-5}$ , determined by bootstrapping.

The benchmarking results of the controlled-phase gate are shown in Fig. 3b. Similar to the single-qubit case, we generate sequences of two-qubit Cliffords to produce a reference curve, and then interleave the controlled-phase gate to extract the fidelity. An example pulse sequence for an  $m = 55$  Clifford sequence is shown in Fig. 3c. We find a controlled-phase gate fidelity of up to  $99.44 \pm 0.05\%$ , consistent with the average error per Clifford (Supplementary Information). We find fidelities of between 99.0% and 99.44% for all four pairs of nearest-neighbour qubits (Supplementary Information). This is a clear demonstration of high-fidelity single- and two-qubit gates in a multi-qubit Josephson quantum processor. The two-qubit gate fidelity compares well with the highest values reported for other mature quantum systems: for nuclear magnetic resonance and ion traps, entangling gate fidelities are as high as 99.5% (ref. 11) and 99.3% (ref. 12). Importantly, we have verified by simulation that the experimentally obtained gate fidelities are at the threshold for surface code quantum error correction (Supplementary Information).

We are optimistic that we can improve on these gate fidelities with modest effort. The controlled-phase gate fidelity is limited by three error mechanisms: decoherence (55% of the total error), control error (24%) and state leakage (21%) (Supplementary Information). Decoherence can be suppressed with enhanced materials and optimized fabrication<sup>26,27</sup>. Imperfections in control arise primarily from reflections and stray inductances in wiring, and can be improved using conventional microwave techniques. Given the adiabatic nature of the controlled-phase gate, two-state leakage can be suppressed by slightly increasing the gate time<sup>10</sup>.

We showcase the modularity of this set of quantum logic gates by constructing a maximally entangled Greenberger–Horne–Zeilinger (GHZ)

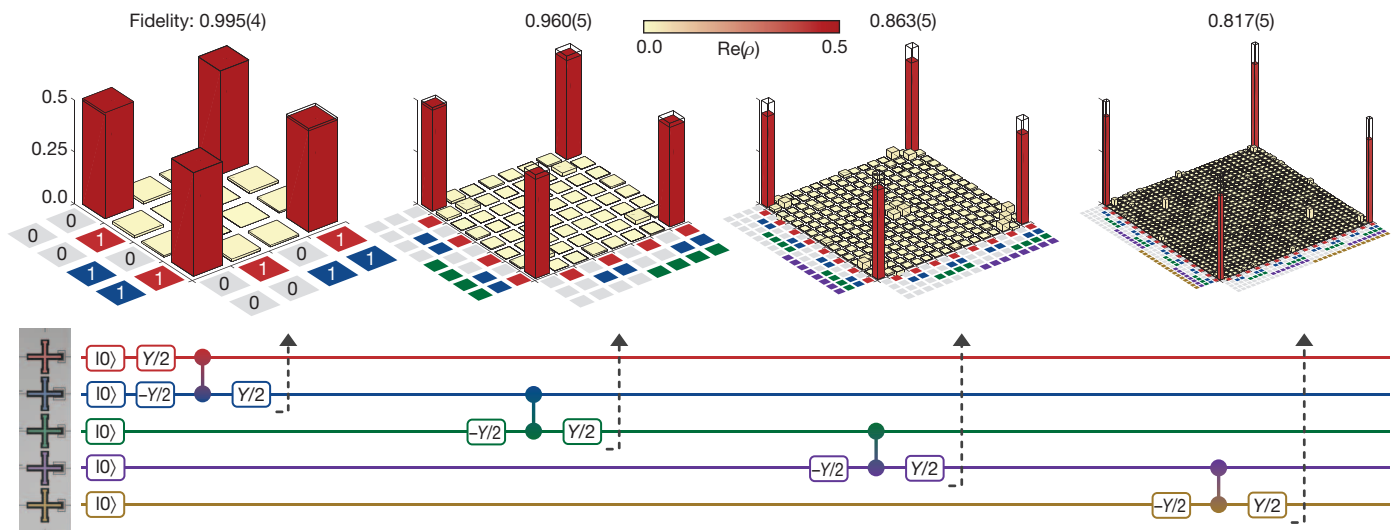




**Figure 3 | Controlled-phase gate physics and randomized benchmarking results.** **a**, We use the  $|1_B 1_A\rangle$  and  $|0_B 2_A\rangle$  avoided level crossing to implement a high-fidelity controlled-phase gate, with the fast adiabatic tuning of qubit A giving a selective  $\pi$  phase change of  $|1_B 1_A\rangle$ . The energy level diagram shows qubit A approaching and leaving the avoided level crossing from above (top, blue dashed line), following a fast (43 ns) yet effectively adiabatic trajectory (bottom, solid blue line). Unwanted state leakage from  $|1_B 1_A\rangle$  to  $|0_B 2_A\rangle$  (red dashed line) is minimized by adjusting the trajectory. **b**, Randomized benchmarking data ( $k = 100$ ) of the controlled-phase gate (CZ) for the qubit pair  $Q_2$  and  $Q_3$ , using the two-qubit Clifford group,  $C_2$  (Supplementary Information). Reference data are in black ( $r_{\text{ref}} = 0.0189$ ); interleaved data are in blue ( $r_{C_2 + CZ} = 0.0244$ ). Dashed lines indicate the thresholds for gate fidelities of 0.98 and 0.99. We find a controlled-phase gate fidelity of  $0.9944 \pm 0.0005$  (uncertainty from bootstrapping). **c**, Coherent microwave (XY) and frequency (Z) control of the quantum state while performing a complex two-qubit algorithm; the sequence contains more than 500 gates, corresponding to the characteristic reference decay of  $m = 55$ , and is more than  $7 \mu\text{s}$  long. The rightmost panel shows an example Clifford from the iSWAP class, comprising single-qubit rotations and two controlled-phase gates (Supplementary Information).

qubits. This algorithm is shown in Fig. 4 (bottom), the state is assembled by entangling one additional qubit at a time. The algorithm is highly sensitive to control error and decoherence on any computational element. We fully characterize the Bell and GHZ states by using quantum state tomography<sup>9</sup>, where quadratic maximum-likelihood estimation is used to extract each density matrix,  $\rho$ , from the measurement data, while satisfying the physical constraints that  $\rho$  be Hermitian, have unit trace and be positive semi-definite (Supplementary Information). The density matrices are plotted in the traditional cityscape style, and show significant elements only at the ideal locations. We find respective state fidelities of  $\text{Tr}(\rho_{\text{ideal}}\rho) = 99.5\% \pm 0.4\%$ ,  $96.0\% \pm 0.5\%$ ,  $86.3\% \pm 0.5\%$  and  $81.7\% \pm 0.5\%$  for the  $N = 2$  Bell state and the  $N = 3, 4$  and  $5$  GHZ states. A GHZ state fidelity of more than 50% satisfies the criterion for genuine entanglement<sup>28</sup>. It is interesting to note that the ratios of off-diagonal to diagonal amplitudes  $|\rho_{|0\rangle^{\otimes N}, |1\rangle^{\otimes N}}|^2 / \rho_{|0\rangle^{\otimes N}, |0\rangle^{\otimes N}} \rho_{|1\rangle^{\otimes N}, |1\rangle^{\otimes N}}$  have the values 0.99, 0.98, 0.99 and 0.99, suggesting that dephasing is small, uncorrelated or both. The five-qubit GHZ state is the largest tomographic measurement of multi-qubit entanglement demonstrated so far in the solid state<sup>8,9</sup>, and has a state fidelity similar to results obtained in ion traps<sup>29</sup>. This demonstrates that complex quantum states can be constructed with high fidelity in a modular fashion, highlighting the potential for more intricate algorithms on this multipurpose quantum processor.

state across all five qubits in our processor (Fig. 4, top). The  $N$ -qubit GHZ state  $|\text{GHZ}\rangle = (|0\rangle^{\otimes N} + |1\rangle^{\otimes N})/\sqrt{2}$  is constructed with single- and two-qubit gates, using simultaneous control and readout of all



**Figure 4 | Quantum state tomography and generation of the GHZ states.** Top: respective real parts of the density matrix  $\rho$  for the  $N = 2$  Bell state and the  $N = 3, 4$  and  $5$  GHZ states, measured by quantum state tomography. Ideal density matrix elements are transparent, with value 0.5 at the four corners.

Bottom: algorithm used to construct the states. See Supplementary Information for  $\text{Im}(\rho)$ , the Pauli operator representation and the full gate sequence, which includes Hahn spin-echo pulses.

We have shown single- and two-qubit gates with fidelities at the fault-tolerant threshold for the surface code in an integrated circuit quantum processor. With this demonstration, Josephson quantum devices are now ready to explore fault-tolerant, multi-qubit computing. Extending the linear array of qubits to larger numbers of qubits is straightforward, and generating a two-dimensional grid of qubits seems to be reduced to the engineering challenge, albeit considerable, of adding more qubits, wiring and readout while maintaining coherence and gate fidelity (Supplementary Information). In a separate experiment, we have demonstrated qubit state measurement with 99% fidelity in 140 ns (ref. 17), with a design that can be seamlessly integrated with this architecture. The combination of high-fidelity logic, a multi-qubit architecture, and fast and accurate qubit readout provides the essential ingredients for a Josephson surface code quantum computer.

## METHODS SUMMARY

We have achieved high-fidelity single-qubit gates with a series of automated experiments that minimize the error in frequency, amplitude and two-state leakage. The qubit frequency is measured by performing a Ramsey experiment. The amplitudes for  $\pi$  and  $\pi/2$  rotations are optimized by concatenating many gates to amplify small rotation errors. The two-state leakage is optimized by minimizing phase error accumulated through repeated  $X/2$  and  $-X/2$  gates; see ref. 21 for details.

The controlled-phase waveform was optimized for phase shift and adiabaticity. We determine the conditional phase through a Ramsey experiment as in Fig. 1c. The qubit trajectory is optimized for adiabaticity when leakage from  $|1_B 1_A\rangle$  to  $|0_B 2_A\rangle$  is minimized. We quantify leakage by initializing the qubits to  $|1_B 1_A\rangle$ , performing a controlled-phase gate and measuring leakage into  $|0_B 2_A\rangle$ .

Parameters for single- and two-qubit gates were fine-tuned or confirmed to be optimal by using Clifford-based randomized benchmarking.

For randomized benchmarking, we generate  $k$  random sequences of Cliffords for each sequence length  $m$ .

Received 26 November 2013; accepted 24 February 2014.

- Vandersypen, L. M. K. *et al.* Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature* **414**, 883–887 (2001).
- Lucero, E. *et al.* Computing prime factors with a Josephson phase qubit quantum processor. *Nature Phys.* **8**, 719–723 (2012).
- Jones, J., Mosca, M. & Hansen, R. Implementation of a quantum search algorithm on a quantum computer. *Nature* **393**, 334–346 (1998).
- Chuang, I. L., Gershenfeld, N. & Kubinec, M. Experimental implementation of fast quantum searching. *Phys. Rev. Lett.* **80**, 3408 (1998).
- Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* Ch. 10 (Cambridge Univ. Press, 2010).
- Fowler, A. G., Mariantoni, M., Martinis, J. M. & Cleland, A. N. Surface codes: towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
- DiCarlo, L. *et al.* Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* **467**, 574–578 (2010).
- Neeley, M. *et al.* Generation of three-qubit entangled states using superconducting phase qubits. *Nature* **467**, 570–573 (2010).
- Martinis, J. & Geller, M. R. Fast adiabatic control of qubits using optimal windowing theory. Preprint at <http://arxiv.org/abs/1402.5467>.
- Ryan, C. A., Laforest, M. & Laflamme, R. Randomized benchmarking of single- and multi-qubit control in liquid-state NMR quantum information processing. *New J. Phys.* **11**, 013034 (2009).
- Benhelm, J., Kirchmair, G., Roos, C. F. & Blatt, R. Towards fault-tolerant quantum computing with trapped ions. *Nature Phys.* **4**, 463–466 (2008).
- Choi, T. *et al.* Optimal quantum control of multi-mode couplings between trapped ion qubits for scalable entanglement. Preprint at <http://arxiv.org/abs/1401.1575> (2014).
- Chow, J. M. *et al.* Implementing a strand of a scalable fault-tolerant quantum computing fabric. Preprint at <http://arxiv.org/abs/1311.6330> (2013).
- Barends, R. *et al.* Coherent Josephson qubit suitable for scalable quantum integrated circuits. *Phys. Rev. Lett.* **111**, 080502 (2013).
- Wallraff, A. *et al.* Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Jeffrey, E. *et al.* Fast scalable state measurement with superconducting qubits. Preprint at <http://arxiv.org/abs/1401.0257> (2014).
- Brown, K. R. *et al.* Single-qubit-gate error below  $10^{-4}$  in a trapped ion. *Phys. Rev. A* **84**, 030303 (2011).
- Córcoles, A. D. *et al.* Process verification of two-qubit quantum gates by randomized benchmarking. *Phys. Rev. A* **87**, 030301 (2013).
- Magesan, E., Gambetta, J. M. & Emerson, J. Scalable and robust randomized benchmarking of quantum processes. *Phys. Rev. Lett.* **106**, 180504 (2011).
- Lucero, E. *et al.* Reduced phase error through optimized control of a superconducting qubit. *Phys. Rev. A* **82**, 042339 (2010).
- Gambetta, J. M. *et al.* Characterization of addressability by simultaneous randomized benchmarking. *Phys. Rev. Lett.* **109**, 240504 (2012).
- DiCarlo, L. *et al.* Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Mariantoni, M. *et al.* Implementing the quantum von Neumann architecture with superconducting circuits. *Science* **334**, 61–65 (2011).
- Strauch, F. W. *et al.* Quantum logic gates for coupled superconducting phase qubits. *Phys. Rev. Lett.* **91**, 167005 (2003).
- Megrant, A. *et al.* Planar superconducting resonators with internal quality factors above one million. *Appl. Phys. Lett.* **100**, 113510 (2012).
- Sendelbach, S., Hover, D., Mück, M. & McDermott, R. Complex inductance, excess noise, and surface magnetism in dc SQUIDS. *Phys. Rev. Lett.* **103**, 117001 (2009).
- Gühne, O. & Seevinck, M. Separability criteria for genuine multiparticle entanglement. *New J. Phys.* **12**, 053002 (2010).
- Monz, T. *et al.* 14-qubit entanglement: creation and coherence. *Phys. Rev. Lett.* **106**, 130506 (2011).
- Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank F. Wilhelm, D. Egger, and J. Baselmans for discussions. We are indebted to E. Lucero for photography of the device. This work was supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office grants W911NF-09-1-0375 and W911NF-10-1-0334. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the US Government. Devices were made at the UC Santa Barbara Nanofabrication Facility, a part of the US NSF-funded National Nanotechnology Infrastructure Network, and at the NanoStructures Cleanroom Facility.

**Author Contributions** R.B. and J.K. designed the sample, performed the experiment and analysed the data. J.K., A.M. and R.B. fabricated the sample. R.B., J.K., J.M.M. and A.N.C. co-wrote the manuscript. A.V. and A.N.K. provided assistance with randomized benchmarking. A.G.F. verified the experimental gate fidelities to be at the surface code threshold. All authors contributed to the fabrication process, experimental set-up and manuscript preparation.

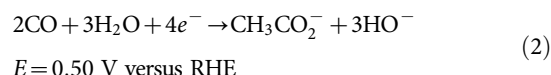
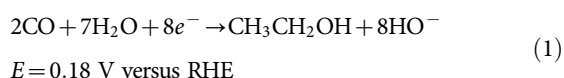
**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B. ([rbarends@physics.ucsb.edu](mailto:rbarends@physics.ucsb.edu)) or J.M.M. ([martinis@physics.ucsb.edu](mailto:martinis@physics.ucsb.edu)).

# Electroreduction of carbon monoxide to liquid fuel on oxide-derived nanocrystalline copper

Christina W. Li<sup>1</sup>, Jim Ciston<sup>2</sup> & Matthew W. Kanan<sup>1</sup>

The electrochemical conversion of CO<sub>2</sub> and H<sub>2</sub>O into liquid fuel is ideal for high-density renewable energy storage and could provide an incentive for CO<sub>2</sub> capture. However, efficient electrocatalysts for reducing CO<sub>2</sub> and its derivatives into a desirable fuel<sup>1–3</sup> are not available at present. Although many catalysts<sup>4–11</sup> can reduce CO<sub>2</sub> to carbon monoxide (CO), liquid fuel synthesis requires that CO is reduced further, using H<sub>2</sub>O as a H<sup>+</sup> source. Copper (Cu) is the only known material with an appreciable CO electroreduction activity, but in bulk form its efficiency and selectivity for liquid fuel are far too low for practical use. In particular, H<sub>2</sub>O reduction to H<sub>2</sub> outcompetes CO reduction on Cu electrodes unless extreme overpotentials are applied, at which point gaseous hydrocarbons are the major CO reduction products<sup>12,13</sup>. Here we show that nanocrystalline Cu prepared from Cu<sub>2</sub>O ('oxide-derived Cu') produces multi-carbon oxygenates (ethanol, acetate and *n*-propanol) with up to 57% Faraday efficiency at modest potentials (–0.25 volts to –0.5 volts versus the reversible hydrogen electrode) in CO-saturated alkaline H<sub>2</sub>O. By comparison, when prepared by traditional vapour condensation, Cu nanoparticles with an average crystallite size similar to that of oxide-derived copper produce nearly exclusive H<sub>2</sub> (96% Faraday efficiency) under identical conditions. Our results demonstrate the ability to change the intrinsic catalytic properties of Cu for this notoriously difficult reaction by growing interconnected nanocrystallites from the constrained environment of an oxide lattice. The selectivity for oxygenates, with ethanol as the major product, demonstrates the feasibility of a two-step conversion of CO<sub>2</sub> to liquid fuel that could be powered by renewable electricity.

Cu electrodes have been thoroughly evaluated in both CO<sub>2</sub> (refs 7, 14, 15) and CO reduction<sup>12,13,16,17</sup> electrolyses. In CO<sub>2</sub>-saturated aqueous solutions, polycrystalline Cu foil produces a mixture of compounds that are dominated by H<sub>2</sub> at low overpotential, by CO and HCO<sub>2</sub><sup>–</sup> at high overpotential and by hydrocarbons and multi-carbon oxygenates at the most extreme potentials<sup>18,19</sup>. When supplied with CO in the absence of CO<sub>2</sub>, Cu produces hydrocarbons and multi-carbon oxygenates, but very negative potentials are still required to promote CO reduction over H<sub>2</sub> evolution<sup>12,13</sup>. Large overpotentials preclude energetically efficient electrolysis and favour hydrocarbons over liquid oxygenates. We recently discovered that oxide-derived Cu (OD-Cu) has much higher selectivity for CO<sub>2</sub> electroreduction over H<sub>2</sub> evolution than does polycrystalline Cu (ref. 20). Because CO reduction activity may be obscured in the presence of a large excess of CO<sub>2</sub>, we questioned whether OD-Cu would reduce CO with greater efficiency than other Cu materials when supplied with CO directly. Here we show that OD-Cu reduces CO with high Faraday efficiency at an overpotential >0.6 V lower than that of polycrystalline Cu foil. This improvement results both from an increase in the intrinsic CO reduction activity and from a decrease in intrinsic H<sub>2</sub>O reduction activity on OD-Cu. The major CO reduction products are ethanol and acetate, which correspond to 8e<sup>–</sup> and 4e<sup>–</sup> reductions with H<sub>2</sub>O as the H<sup>+</sup> source (equations (1) and (2), where *E* is the equilibrium potential).



By comparing OD-Cu to electrodes comprised of commercial Cu nanoparticles, we show that CO reduction activity is not a consequence of nanocrystallite size or morphology; instead, we propose that grain boundaries participate in the catalysis.

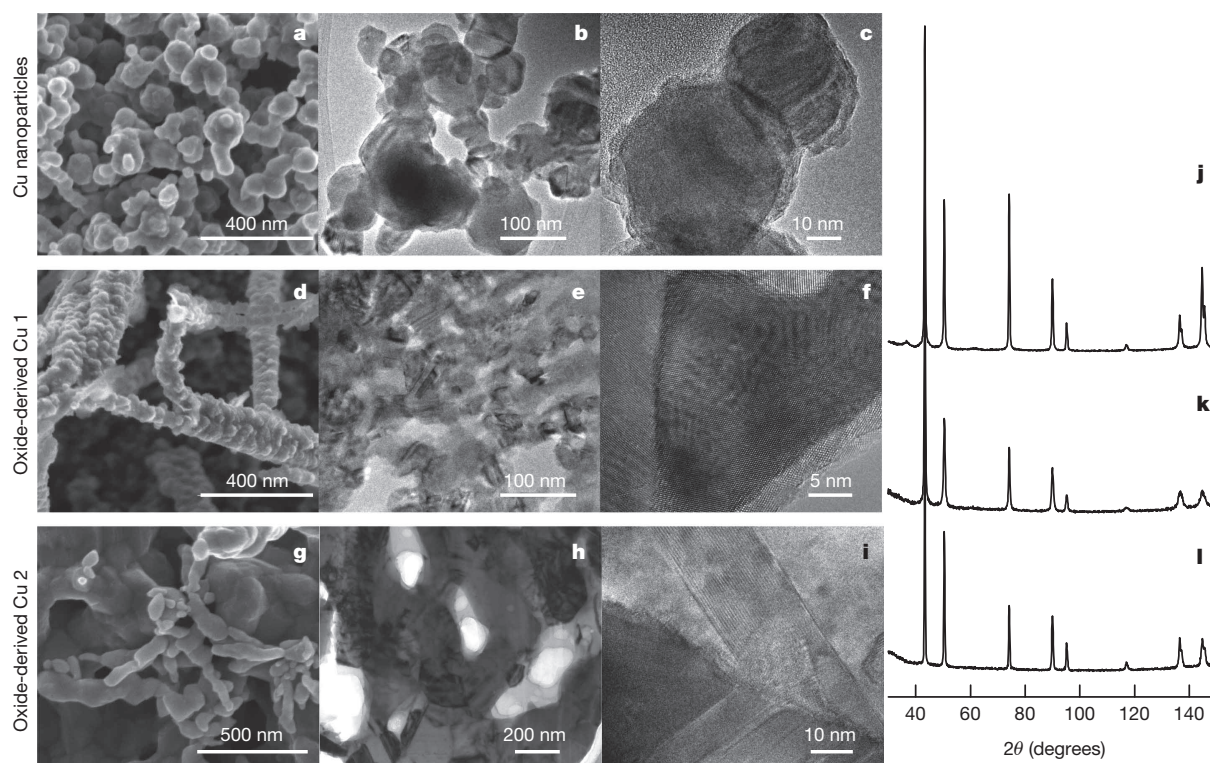
OD-Cu electrodes were prepared by annealing polycrystalline Cu foil in air at 500 °C to grow a thick Cu<sub>2</sub>O layer on the surface and subsequently reducing this oxide to form Cu<sup>0</sup> nanocrystallites. We used two reduction methods to vary the properties of the material. To obtain OD-Cu 1, the Cu<sub>2</sub>O precursor was reduced electrochemically in aqueous solution at ambient temperature. This procedure had previously been found to yield nanocrystalline Cu with maximal selectivity for CO<sub>2</sub> reduction<sup>20</sup>. To obtain OD-Cu 2, the Cu<sub>2</sub>O precursor was reduced with H<sub>2</sub> at 130 °C. The OD-Cu electrodes were compared to commercial Cu nanoparticles that had been synthesized by vapour condensation. Cu nanoparticle electrodes were prepared by drop-casting suspensions of the nanoparticles and fluorinated binder onto metal substrates.

All three types of electrodes exhibited similar morphologies when imaged by scanning electron microscopy (SEM) (Fig. 1a, d, g). Particle sizes ranged from about 30 nm to 100 nm, and the particles formed aggregated films. Low-resolution transmission electron microscopy (TEM) confirmed the range of particle sizes observed via SEM (Fig. 1b, e, h). High-resolution TEM of OD-Cu 1 and OD-Cu 2 showed interconnected nanocrystalline networks with distinct grain boundaries between nanocrystallites (Fig. 1f, i). In contrast, the Cu nanoparticle electrode appeared as an aggregation of overlapping particles (Fig. 1c). Differential capacitance measurements yielded electrochemical roughness factors of 135, 48 and 26 for OD-Cu 1, OD-Cu 2 and the Cu nanoparticle electrode, respectively (Extended Data Fig. 5 and Extended Data Table 2). X-ray photoelectron spectroscopy indicated that the surfaces were comprised of both Cu<sup>0</sup> and Cu<sub>2</sub>O, the latter resulting from air exposure during the X-ray photoelectron spectroscopy sample preparation (Extended Data Figs 1, 2 and 3). Grazing incidence X-ray diffraction patterns for all of the electrodes exhibited peaks at the expected positions for an ideal Cu lattice, indicating no uniform expansion or compression of the unit cell (Fig. 1j, k, l). The patterns differed substantially in their peak widths, however. Using the Scherrer equation, both OD-Cu 2 and the Cu nanoparticle electrode exhibited average crystallite sizes ⟨*D*⟩ > 100 nm, while OD-Cu 1 had a smaller ⟨*D*⟩ of 31 nm. Additional high-resolution grazing incidence X-ray diffraction patterns were collected for OD-Cu 1 and OD-Cu 2 with synchrotron radiation to determine the amount of microstrain in each sample using Williamson–Hall analysis. This analysis yielded microstrain values of 0.2% and 0.05% for OD-Cu 1 and OD-Cu 2, respectively (Extended Data Fig. 4).

CO electroreduction activity was measured under steady-state conditions by performing constant-potential electrolyses for multiple hours in 0.1 M KOH saturated with 1 atm of CO at ambient temperature. The concentration of CO in solution is only 1 mM under these conditions. Solution-phase and gas-phase products were quantified by nuclear magnetic

<sup>1</sup>Department of Chemistry, Stanford University, Stanford 94305, California. <sup>2</sup>National Center for Electron Microscopy, Lawrence Berkeley National Laboratory, Berkeley 94720, California.





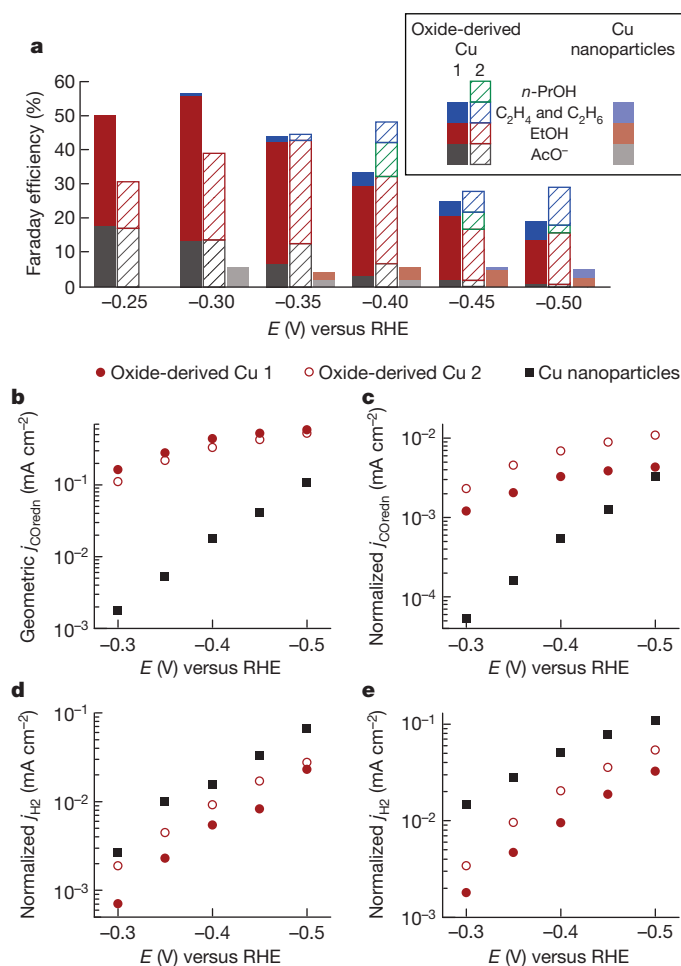
**Figure 1 | Physical characterization of Cu nanoparticle and OD-Cu electrodes.** Top row, the Cu nanoparticle electrode. Middle row, the OD-Cu 1 electrode. Bottom row, the OD-Cu 2 electrode. **a, d, g**, SEM images. **b, e, h**, Low-magnification TEM images. **c, f, i**, High-resolution TEM images. **j, k, l**, Grazing incidence X-ray diffraction patterns.

resonance (NMR) spectroscopy and gas chromatography, respectively. Electrolyses were performed at a range of applied potentials  $E$  between  $-0.25$  V and  $-0.5$  V versus the reversible hydrogen electrode (RHE). The modest reducing potential and low CO concentration provides a stringent test of a catalyst's selectivity for CO versus  $H_2O$  reduction. Under these conditions, polycrystalline Cu foil exhibited low geometric current densities and  $H_2$  was the only detectable product. These results are consistent with previous results in KOH, in which much more negative potentials ( $-0.7$  V versus RHE) and a high flow rate of CO through the solution were required to attain 22% total Faraday efficiency for CO reduction, with only 7% Faraday efficiency for oxygenates<sup>13</sup>. A maximum CO reduction Faraday efficiency of 65% was reported for polycrystalline Cu foil at  $-0.9$  V versus RHE in a mixed KCl and  $KHCO_3$  electrolyte, although only 10% Faraday efficiency corresponded to the formation of oxygenates<sup>12</sup>.

Larger geometric current densities were obtained with OD-Cu and Cu nanoparticle electrodes as a result of their roughness factors, which facilitated product quantification and comparison of their selectivities (Fig. 2 and Extended Data Table 1). For the Cu nanoparticle electrode, the Faraday efficiency for  $H_2$  was  $\geq 94\%$  at all potentials examined; the small remainder of the current corresponded to acetate, ethanol and ethylene formation. The high selectivity for  $H_2O$  reduction on this electrode mimics that of a Cu foil electrode. In contrast, OD-Cu electrodes had a much greater propensity for CO reduction. OD-Cu 1 attained a total CO reduction Faraday efficiency of 57% at  $-0.3$  V, whereas OD-Cu 2 attained 48% at  $-0.4$  V. These values declined at more negative potentials, but this decline was simply the result of the catalysts reaching the mass-transport-limited current density for CO reduction (see below). For both materials, ethanol and acetate were the nearly exclusive CO reduction products at the more positive potentials, whereas ethylene production emerged at more negative potentials. OD-Cu 2 also produced substantial amounts of propanol, which was only observed for OD-Cu 1 under higher CO pressure (see below). In addition to the reduction products, OD-Cu electrodes also produced small amounts of formate, the product of CO hydration.

The higher Faraday efficiency for CO reduction on OD-Cu electrodes compared to a Cu nanoparticle electrode reflected higher intrinsic CO reduction activity. Figure 2b and c shows the geometric and surface-area-normalized current densities for CO reduction ( $j_{COredn}$ ) versus  $E$  for the three electrodes. The difference in intrinsic activity between two electrodes corresponds to the ratio of their normalized current densities in the absence of mass-transport contributions to the kinetics. The rate of CO transport from bulk solution through the diffusion layer to the electrode surface determines the maximum attainable geometric  $j_{COredn}$  (ref. 21). The Cu nanoparticle electrode did not reach a mass-transport limit between  $-0.3$  V and  $-0.5$  V, as shown by the exponential increase in  $j_{COredn}$  across this potential range. Its Tafel slope was  $112$  mV dec<sup>-1</sup> (where 'dec' means decade, or one order of magnitude). On account of their higher activity, the geometric  $j_{COredn}$  on OD-Cu electrodes quickly reached a plateau value of  $0.5$  mA cm<sup>-2</sup> as the overpotential was increased (Fig. 2b), indicating that this value is approximately the mass transport limit with only 1 mM CO in solution. At  $-0.3$  V, before the onset of the mass transport contributions, normalized  $j_{COredn}$  was 23-fold higher for OD-Cu 1 and 43-fold higher for OD-Cu 2 than for the Cu nanoparticle electrode (Fig. 2c). To determine the dependence of  $j_{COredn}$  on  $E$  for OD-Cu electrodes in the absence of mass-transport effects, additional electrolyses were performed in the range  $-0.25$  V  $> E > -0.35$  V (Extended Data Fig. 8). A Tafel slope of approximately  $113$  mV dec<sup>-1</sup> was obtained.

The electrodes also differed in their intrinsic activities for  $H_2O$  reduction. Figure 2d and e shows the normalized current density for  $H_2O$  reduction ( $j_{H2}$ ) in CO-saturated and  $N_2$ -saturated 0.1 M KOH, respectively. All three electrodes exhibited an  $H_2O$  reduction activity 2-fold to 5-fold lower in the presence of CO than in the presence of  $N_2$ , indicating some poisoning of the sites for  $H_2O$  reduction by CO binding<sup>22</sup>. In CO-saturated electrolyte, normalized  $j_{H2}$  was similar for the Cu nanoparticle electrode and OD-Cu 2, but was 2-fold to 6-fold lower for OD-Cu 1. The suppression of  $H_2O$  reduction contributes to the higher Faraday efficiency for CO reduction on OD-Cu 1. In  $N_2$ -saturated electrolyte,  $j_{H2}$  was suppressed 4–8-fold on OD-Cu 1 and 2–6-fold on OD-Cu 2

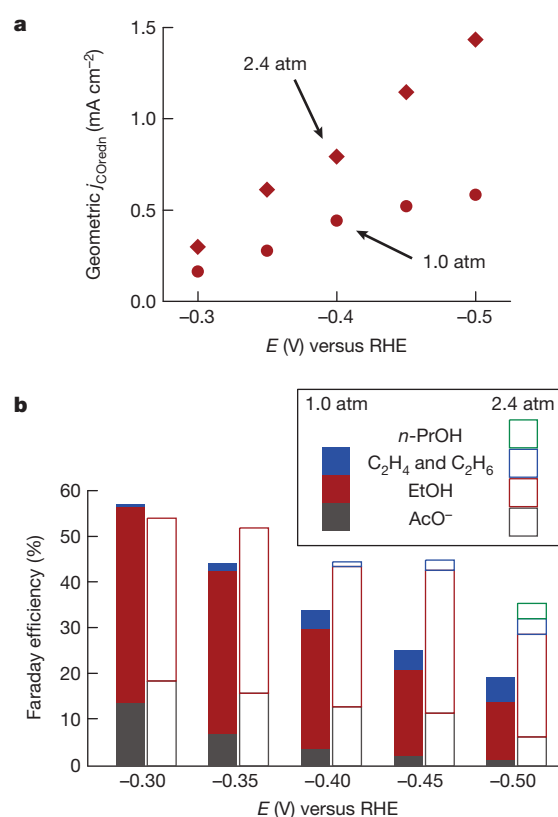


**Figure 2 | Comparison between OD-Cu 1, OD-Cu 2 and Cu nanoparticle electrodes in electrolyses performed in 0.1 M KOH saturated with 1 atm CO at ambient temperature.** **a**, Faraday efficiencies for CO reduction products ethanol (EtOH), acetate (AcO<sup>-</sup>), *n*-propanol (*n*-PrOH), ethylene (C<sub>2</sub>H<sub>4</sub>) and ethane (C<sub>2</sub>H<sub>6</sub>) at selected potentials versus the reversible hydrogen electrode. The remaining Faraday efficiency in all cases corresponds to H<sub>2</sub> evolution. The current density for the Cu nanoparticle electrode at -0.25 V was too low for product quantification. **b**, Geometric current densities for CO reduction. **c**, Surface-area-normalized current densities for CO reduction. **d**, Surface-area-normalized current density for H<sub>2</sub> evolution under 1 atm of CO. **e**, Surface-area-normalized current density for H<sub>2</sub> evolution under 1 atm of N<sub>2</sub>.

compared to the Cu nanoparticle electrode. The similar Tafel slopes for H<sub>2</sub> evolution suggest that OD-Cu electrodes have a lower density of active sites for this reaction.

To confirm that the geometric  $j_{\text{COredn}}$  at 1 atm CO is limited by the kinetics of CO mass transport, OD-Cu 1 was evaluated in an electrolysis performed in 0.1 M KOH saturated with 2.4 atm of CO. At potentials between -0.3 V and -0.5 V, geometric  $j_{\text{COredn}}$  was 1.8–2.4-fold higher under these conditions (Fig. 3a). The Faraday efficiency for CO reduction at  $E < -0.3$  V was also substantially improved (Fig. 3b). These results indicate that practical current densities may be possible in this moderate overpotential regime by further increasing CO transport to the catalyst.

The electrolysis data provide some insight into the mechanism of CO reduction on OD-Cu. The  $j_{\text{COredn}}$  Tafel slope is consistent with a rate-limiting initial  $1e^-$  transfer to CO. The absence of C1 products indicates that C–C coupling is rapid once the process of CO reduction initiates<sup>23</sup> or that the initial  $e^-$  transfer is coupled to C–C bond formation between a surface-bound CO and a CO from solution<sup>16</sup>. The latter would account for increased  $j_{\text{COredn}}$  at increased pressure if the active sites of OD-Cu were saturated with CO at 1 atm (refs 24, 25). The formation of acetate probably arises from attack of HO<sup>-</sup> on a surface-bound ketene or



**Figure 3 | Comparison of CO reduction in 0.1 M KOH saturated with 1 atm of CO versus 2.4 atm of CO.** **a**, Geometric current densities for CO reduction. **b**, Faraday efficiencies for CO reduction products at selected potentials.

other carbonyl-containing intermediate after C–C bond formation has occurred. This contention is supported by the observation of substantially increased acetate formation when the electrolysis is performed in 1 M KOH (Extended Data Fig. 9).

The results above implicate the participation of grain boundary surfaces in the CO reduction catalysis on OD-Cu. Although OD-Cu 1 and OD-Cu 2 differ in their bulk properties, both are distinguished from Cu nanoparticle electrodes by having a high density of grain boundaries as a consequence of their formation from a solid phase. Given that grain boundaries can support surface structures that are not stable on an individual nanocrystal<sup>26</sup>, the enhanced CO reduction activity on OD-Cu electrodes relative to Cu nanoparticle electrodes is consistent with the presence of highly active sites on their grain boundary surfaces. The difference in their product distribution could reflect differences in grain boundary structures resulting from electrochemical versus H<sub>2</sub> reduction of Cu<sub>2</sub>O. The results suggest that engineering the grain boundaries by altering the oxide reduction method or by using an alternative preparation of nanocrystalline materials could yield catalysts with higher activity and selectivity for long-chain products. The participation of grain boundary surfaces in catalysis has previously been proposed to explain correlations between grain boundary density and methanol electrooxidation activity with Pt nanoparticles<sup>27</sup> and PtRu alloy nanoparticles<sup>28</sup>.

Catalytic processes have been optimized over the past century to reduce CO with H<sub>2</sub> to alkanes, methanol or higher alcohols. Although many of these processes are efficient, they all require high temperatures and pressures (typically 250–400 °C and 50–150 atm) and large-scale reactors that are poorly matched to diffuse renewable energy sources<sup>29,30</sup>. Moreover, H<sub>2</sub> is almost exclusively derived from fossil fuel sources at present. Electrolytic syntheses enabled by the catalysis described here could provide a more direct, versatile and energetically efficient route to distributed liquid fuel production powered by renewable energy.

## METHODS SUMMARY

CO reduction electrolyses were performed in CO-saturated 0.1 M KOH electrolyte in a two-compartment electrochemical cell. The headspace of the cathodic chamber was continuously purged with CO into the sampling loop of a gas chromatograph to enable periodic quantification of the gas-phase products. The solution-phase products were quantified by NMR analysis of the electrolyte at the conclusion of the electrolyses. OD-Cu electrodes were prepared by electropolishing pieces of polycrystalline Cu foil (99.9999%) in 85% phosphoric acid and subsequently annealing the electrodes in air at 500 °C for 1 h to form a thick Cu<sub>2</sub>O layer. The active OD-Cu 1 layer was formed *in situ* at the beginning of CO reduction electrolysis, requiring 10–12 C cm<sup>-2</sup> of charge to fully reduce the Cu<sub>2</sub>O material. OD-Cu 2 was formed by reduction of the Cu<sub>2</sub>O layer in flowing H<sub>2</sub> at 130 °C for 2 h.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 June 2013; accepted 27 January 2014.

Published online 9 April 2014.

- Appel, A. M. *et al.* Frontiers, opportunities, and challenges in biochemical and chemical catalysis of CO<sub>2</sub> fixation. *Chem. Rev.* **113**, 6621–6658 (2013).
- Jhong, H. R., Ma, S. & Kenis, P. J. A. Electrochemical conversion of CO<sub>2</sub> to useful chemicals: current status, remaining challenges, and future opportunities. *Curr. Opin. Chem. Eng.* **2**, 191–199 (2013).
- Cole, E. B. & Bocarsly, A. B. in *Carbon Dioxide as Chemical Feedstock* (ed. Aresta, M.) 291–316 (Wiley, 2010).
- Costentin, C., Robert, M. & Saveant, J. M. Catalysis of the electrochemical reduction of carbon dioxide. *Chem. Soc. Rev.* **42**, 2423–2436 (2013).
- Benson, E. E., Kubiak, C. P., Sathrum, A. J. & Smieja, J. M. Electrocatalytic and homogeneous approaches to conversion of CO<sub>2</sub> to liquid fuels. *Chem. Soc. Rev.* **38**, 89–99 (2009).
- Costentin, C., Drouet, S., Robert, M. & Saveant, J. M. A local proton source enhances CO<sub>2</sub> electroreduction to CO by a molecular Fe catalyst. *Science* **338**, 90–94 (2012).
- Hori, Y. in *Modern Aspects of Electrochemistry* Vol. 42 (eds Vayenas, C. G., White, R. E. & Gamboa-Aldeco, M. E.) 89–189 (Springer, 2008).
- Chen, Y., Li, C. W. & Kanan, M. W. Aqueous CO<sub>2</sub> reduction at very low overpotential on oxide-derived Au nanoparticles. *J. Am. Chem. Soc.* **134**, 19969–19972 (2012).
- Tornow, C. E., Thorson, M. R., Ma, S., Gewirth, A. A. & Kenis, P. J. A. Nitrogen-based catalysts for the electrochemical reduction of CO<sub>2</sub> to CO. *J. Am. Chem. Soc.* **134**, 19520–19523 (2012).
- DiMeglio, J. L. & Rosenthal, J. Selective conversion of CO<sub>2</sub> to CO with high efficiency using an inexpensive bismuth-based electrocatalyst. *J. Am. Chem. Soc.* **135**, 8798–8801 (2013).
- Ebbesen, S. D. & Mogensen, M. Electrolysis of carbon dioxide in solid oxide electrolysis cells. *J. Power Sources* **193**, 349–358 (2009).
- Hori, Y., Takahashi, R., Yoshinami, Y. & Murata, A. Electrochemical reduction of CO at a copper electrode. *J. Phys. Chem. B* **101**, 7075–7081 (1997).
- Hori, Y., Murata, A., Takahashi, R. & Suzuki, S. Electroreduction of CO to CH<sub>4</sub> and C<sub>2</sub>H<sub>4</sub> at a copper electrode in aqueous solutions at ambient temperature and pressure. *J. Am. Chem. Soc.* **109**, 5022–5023 (1987).
- Gattrell, M., Gupta, N. & Co, A. A review of the aqueous electrochemical reduction of CO<sub>2</sub> to hydrocarbons at copper. *J. Electroanal. Chem.* **594**, 1–19 (2006).
- Hori, Y., Takahashi, I., Koga, O. & Hoshi, N. Selective formation of C<sub>2</sub> compounds from electrochemical reduction of CO<sub>2</sub> at a series of copper single crystal electrodes. *J. Phys. Chem. B* **106**, 15–17 (2002).
- Calle-Vallejo, F. & Koper, M. T. M. Theoretical considerations on the electroreduction of CO to C<sub>2</sub> species on Cu(100) electrodes. *Angew. Chem. Int. Ed.* **52**, 7282–7285 (2013).
- Schouten, K. J. P., Qin, Z. S., Gallent, E. P. & Koper, M. T. M. Two pathways for the formation of ethylene in CO reduction on single-crystal copper electrodes. *J. Am. Chem. Soc.* **134**, 9864–9867 (2012).
- Kuhl, K. P., Cave, E. R., Abram, D. N. & Jaramillo, T. F. New insights into the electrochemical reduction of carbon dioxide on metallic copper surfaces. *Energy Environ. Sci.* **5**, 7050–7059 (2012).
- Hori, Y., Murata, A. & Takahashi, R. Formation of hydrocarbons in the electrochemical reduction of carbon dioxide at a copper electrode in aqueous solution. *J. Chem. Soc. Faraday Trans. 1* **85**, 2309–2326 (1989).
- Li, C. W. & Kanan, M. W. CO<sub>2</sub> reduction at low overpotential on Cu electrodes resulting from the reduction of thick Cu<sub>2</sub>O films. *J. Am. Chem. Soc.* **134**, 7231–7234 (2012).
- Gileadi, E. *Electrode Kinetics for Chemists, Engineers, and Materials Scientists* Ch. 1 1–8 (Wiley, 1993).
- Hori, Y., Murata, A. & Yoshinami, Y. Adsorption of CO, intermediately formed in electrochemical reduction of CO<sub>2</sub>, at a copper electrode. *J. Chem. Soc. Faraday Trans. 87*, 125–128 (1991).
- Montoya, J. H., Peterson, A. A. & Nørskov, J. K. Insights into C–C coupling in CO<sub>2</sub> electroreduction on copper electrodes. *ChemCatChem* **5**, 737–742 (2013).
- Koga, O. *et al.* Infrared spectroscopic and voltammetric study of adsorbed CO on stepped surfaces of copper monocrystalline electrodes. *Electrochim. Acta* **50**, 2475–2485 (2005).
- Shaw, S. K. *et al.* Role of axially coordinated surface sites for electrochemically controlled carbon monoxide adsorption on single crystal copper electrodes. *Phys. Chem. Chem. Phys.* **13**, 5242–5251 (2011).
- Radetic, T., Lancon, F. & Dahmen, U. Chevron defect at the intersection of grain boundaries with free surfaces in Au. *Phys. Rev. Lett.* **89**, 085502 (2002).
- Wang, S. Y., Jiang, S. P., White, T. J., Guo, J. & Wang, X. Electrocatalytic activity and interconnectivity of Pt nanoparticles on multiwalled carbon nanotubes for fuel cells. *J. Phys. Chem. C* **113**, 18935–18945 (2009).
- Gavrilov, A. N. *et al.* On the influence of the metal loading on the structure of carbon-supported PtRu catalysts and their electrocatalytic activities in CO and methanol electrooxidation. *Phys. Chem. Chem. Phys.* **9**, 5476–5489 (2007).
- Mills, G. A. Status and future opportunities for conversion of synthesis gas to liquid fuels. *Fuel* **73**, 1243–1279 (1994).
- Gupta, M., Smith, M. L. & Spivey, J. J. Heterogeneous catalytic conversion of dry syngas to ethanol and higher alcohols on Cu-based catalysts. *ACS Catal.* **1**, 641–656 (2011).

**Acknowledgements** We thank Stanford University and the NSF (CHE-1266401) for support of this work. C.W.L. gratefully acknowledges an NSF Predoctoral Fellowship. A portion of this work was performed at NCEM, which is supported by the Office of Science, Office of Basic Energy Sciences of the US Department of Energy under contract number DE-AC02-05CH11231. We thank M. Toney and B. Shyam for assistance with grazing incidence X-ray diffraction performed at SSRL, a national user facility operated by Stanford University on behalf of the Office of Basic Energy Sciences of the US Department of Energy.

**Author Contributions** C.W.L. and M.W.K. designed the experiments. C.W.L. prepared and characterized all electrodes and performed all electrochemical experiments; J.C. obtained all TEM images; C.W.L. and M.W.K. wrote the manuscript. All authors contributed to the overall scientific interpretation and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.W.K. ([mkanan@stanford.edu](mailto:mkanan@stanford.edu)).



# Hybrid shallow on-axis and deep off-axis hydrothermal circulation at fast-spreading ridges

Jörg Hasenclever<sup>1</sup>, Sonja Theissen-Krah<sup>2</sup>, Lars H. Rüpke<sup>1</sup>, Jason P. Morgan<sup>3</sup>, Karthik Iyer<sup>1</sup>, Sven Petersen<sup>1</sup> & Colin W. Devey<sup>1</sup>

Hydrothermal flow at oceanic spreading centres accounts for about ten per cent of all heat flux in the oceans<sup>1,2</sup> and controls the thermal structure of young oceanic plates. It also influences ocean and crustal chemistry, provides a basis for chemosynthetic ecosystems, and has formed massive sulphide ore deposits throughout Earth's history. Despite this, how and under what conditions heat is extracted, in particular from the lower crust, remains largely unclear. Here we present high-resolution, whole-crust, two- and three-dimensional simulations of hydrothermal flow beneath fast-spreading ridges that predict the existence of two interacting flow components, controlled by different physical mechanisms, that merge above the melt lens to feed ridge-centred vent sites. Shallow on-axis flow structures develop owing to the thermodynamic properties of water, whereas deeper off-axis flow is strongly shaped by crustal permeability, particularly the brittle-ductile transition. About 60 per cent of the discharging fluid mass is replenished on-axis by warm (up to 300 degrees Celsius) recharge flow surrounding the hot thermal plumes, and the remaining 40 per cent or so occurs as colder and broader recharge up to several kilometres away from the axis that feeds hot (500–700 degrees Celsius) deep-rooted off-axis flow towards the ridge. Despite its lower contribution to the total mass flux, this deep off-axis flow carries about 70 per cent of the thermal energy released at the ridge axis. This combination of two flow components explains the seismically determined thermal structure of the crust and reconciles previously incompatible models favouring either shallower on-axis<sup>3–5</sup> or deeper off-axis hydrothermal circulation<sup>6–8</sup>.

The fast-spreading East Pacific Rise (EPR) at 9° N is one of the best-studied ridge sections worldwide. Multichannel seismic studies between 9° N and 13° N image a nearly continuous sub-axial melt lens at depths of about 1.2–2.4 km (ref. 9). A tomographic study<sup>10</sup> at 9° 30' N shows a narrow P-wave anomaly below the melt lens that is best modelled as a 4–5-km-wide, high-temperature region extending through the whole crust and only widening at depths below the crust–mantle boundary (Fig. 1a). Isotherms are steep near this 'hot slot' but become nearly horizontal further off-axis, consistent with field observations from the Oman ophiolite<sup>7,11</sup>. Axis-perpendicular two-dimensional (2D) hydrothermal flow models<sup>12,13</sup> show that this implies hydrothermal convection through the entire crust<sup>14</sup>, with hydrothermal fluids probably migrating parallel to the steep isotherms and thereby altering the lower gabbros at some distance away from the ridge axis<sup>15</sup>. This is in contrast to the conclusion, drawn from microearthquake distributions<sup>5</sup> at EPR 9° N, that hydrothermal cells develop predominantly above the melt lens—a concept apparently supported by vent fluid chemistry studies<sup>16</sup> and previous three-dimensional (3D) simulations of hydrothermal flow at fast-spreading ridges<sup>4,17</sup>.

To investigate how whole-crustal cooling can be reconciled with supra-melt lens fluid circulation, we developed a 3D model that resolves hydrothermal flow throughout the entire thickness of the oceanic crust at a fast-spreading ridge. We obtain the permeability structure for this 3D model from 2D numerical experiments in which we simultaneously solve for crustal accretion processes and cooling by hydrothermal flow<sup>18</sup>. This

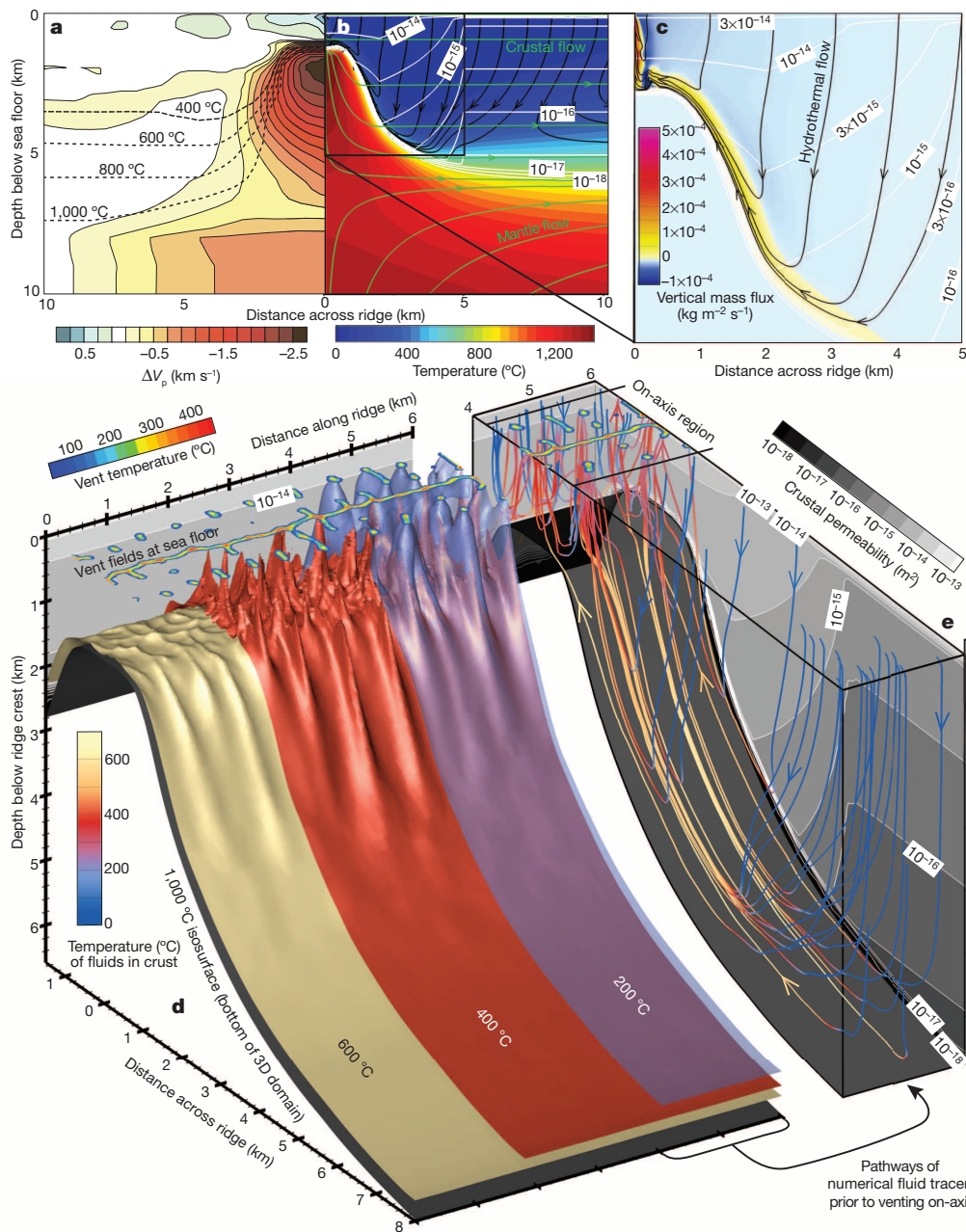
2D model predicts the thermal structure of the young crust and the depth of the melt lens for a given permeability field and a full spreading rate of 11 cm yr<sup>−1</sup> (see ref. 18 and Methods). Permeability decreases exponentially with depth and linearly with distance from the ridge axis and was varied until (1) the predicted crustal temperature field matches the one inferred from seismic tomography<sup>10</sup> and (2) the melt lens (corresponding to the 1,200 °C isotherm) is located at the observed depth<sup>9</sup> of 1,400 m. Key features of this parameterized permeability field (see Methods and Extended Data Fig. 1) are relatively high permeabilities adjacent to the margins of the hot slot, mimicking the effects of a thermal cracking front<sup>19</sup>. This is consistent with microcracks found in lower gabbros<sup>7</sup> and in peridotites a few hundred metres below the crust–mantle boundary<sup>20</sup> of the Oman ophiolite. An additional decrease in permeability beyond a distance to the ridge of 5 km is required to reproduce the seismically inferred nearly horizontal isotherms<sup>10</sup> in this region and to be consistent with the low permeabilities inferred from porosity logs from drill holes in more mature oceanic crust<sup>21</sup>. The brittle–ductile transition in the lower crust<sup>22</sup> is accounted for by decreasing permeability to 10<sup>−18</sup> m<sup>2</sup> between 600 °C and 800 °C, so that no significant fluid flow occurs through rock hotter than 800 °C. The top of the model has a 250-m-thick pillow basalt layer<sup>23</sup> in which permeability increases to 10<sup>−13</sup> m<sup>2</sup> at the sea floor.

The 2D results (Fig. 1b) show the thermal evolution of the oceanic crust after 500,000 years of simulation time in response to the interplay of crustal accretion processes and hydrothermal cooling. The hydrothermal flow patterns predicted by the 2D model (Fig. 1c) hint at two coexisting flow components: (1) on-axis convection above the melt lens with strong adjacent recharge flow and (2) much broader and deeper off-axis circulation. Both flow components merge above the melt lens and supply the axial discharge.

Using a 3D hydrothermal flow model (see Methods) we constrain the relative importance of on- versus off-axis flow and explore the detailed flow structures developing over shorter timescales of a few thousand years. Whereas previous 3D<sup>4,17</sup> and across-axis 2D<sup>12,13,18</sup> models had focused either on supra-melt lens or ridge-perpendicular flow, we use a whole-crust 3D finite-element model that resolves both flow components simultaneously. Fluid properties are determined from the pure water IAPS-84 equation of state, so that the fluid is in the supercritical single-phase regime for the hydrostatic pressures prevailing at the sea-floor. The 1,000 °C isotherm derived from the 2D results at steady state is used as a constant-temperature lower boundary for the 3D domain, whereas the top boundary follows a smoothed-average bathymetry of the EPR at 9° N. Fluids are free to enter (at a temperature of 2 °C) or to leave the domain through the top boundary only. Sidewalls are impermeable and insulating. This set-up results in a 16 km × 6 km × 6.5 km-sized computational domain that is meshed with about 25 million tetrahedral elements (Extended Data Fig. 2). Each 3D simulation ran on a cluster with 32 processors for about 3 weeks.

Our 3D results show the formation of a single hybrid hydrothermal system (Fig. 1d, e). On-axis upflow organizes into a network of closely interconnected thermal plumes that feed high-temperature (up to 410 °C)

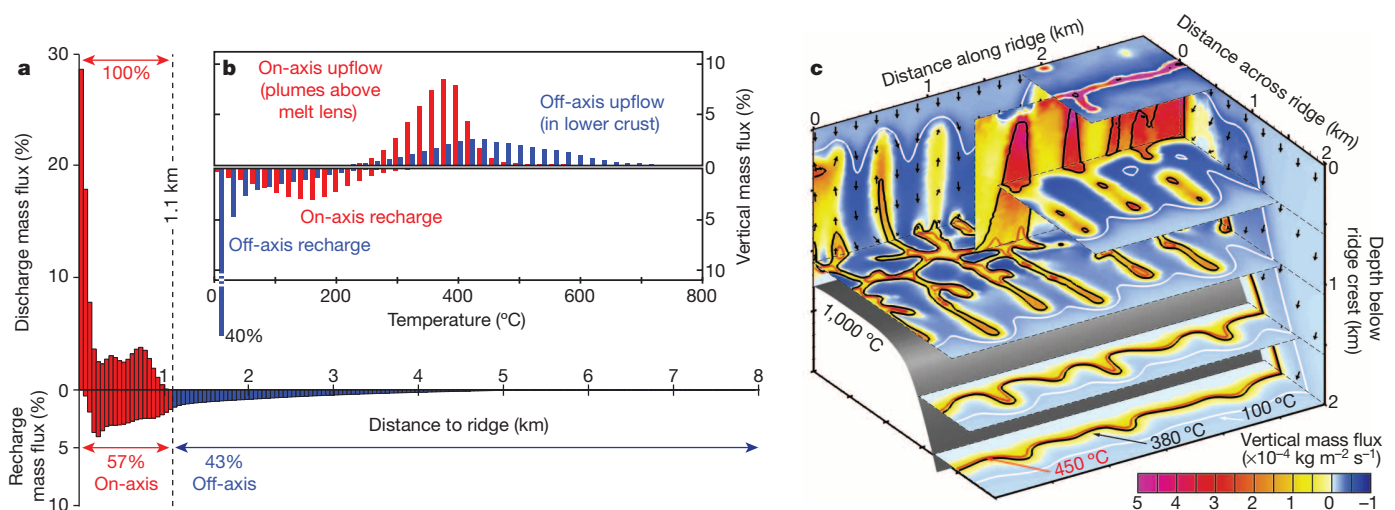
<sup>1</sup>GEOMAR, Helmholtz Centre for Ocean Research Kiel, Wischhofstraße 1-3, 24148 Kiel, Germany. <sup>2</sup>Department of Geosciences and Centre for Earth Evolution and Dynamics (CEED), University of Oslo, PO Box 1048, Blindern, 0316 Oslo, Norway. <sup>3</sup>Department of Earth Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK.



**Figure 1 | Results of our 2D and 3D numerical experiments.** **a**, Seismic and inferred thermal structure beneath the EPR (re-drawn after plate 1 and figure 10 in ref. 10). **b**, Quasi-steady-state temperature field of a best-fitting 2D simulation using our coupled crustal accretion/solid flow (green streamlines) and hydrothermal convection (black streamlines) model. The permeability field is shown as white contours. **c**, Vertical fluid mass flux shows transient narrow on-axis and steady broad off-axis circulation; positive values correspond to upward fluid flow. **d**, Our 3D model resolves hydrothermal convection in a 6-km-long and 16-km-wide modelling domain above the 1,000 °C isotherm (black bottom surface) obtained from the 2D results. Thermal evolution and vent field locations are shown after 3,500 years of simulation time. **e**, Flow paths of fluid tracers, colour-coded for temperature in a subsection of the 3D domain.

axial vent fields with an along-axis spacing between 500 m and 1,000 m. Partial ‘on-axis’ recharge surrounds the thermal plumes and does not penetrate deeper than the depth of the melt lens (see Fig. 1e), consistent with earlier findings by ref. 4. In contrast, ‘off-axis’ recharge flow occurs over a broad area and feeds very hot (500 °C–700 °C) ridge-perpendicular flow upslope the flanks of the hot slot, showing no significant ridge-parallel motion until the melt lens depth is reached, where the fluid starts to focus into tubes with a spacing similar to that of the on-axis plumes (Fig. 1d, e). Near the top of the melt lens both flow components merge before feeding the axial vent fields, where fluids are released at a rate of 80–90 kg s<sup>-1</sup> per kilometre ridge axis. A quantification of the fluid mass fluxes as a function of distance to the ridge (Fig. 2a) shows that, on average, about 60% of the entire recharge flow occurs on-axis (within 1.1 km of the ridge axis), while about 40% of the venting fluids originate from distances greater than 1.1 km from the ridge axis. This ratio also holds for simulations with along-ridge variations in melt lens depth (see Extended Data Figs 4 and 5).

An analysis of vertical mass fluxes versus temperature (Fig. 2b and Extended Data Fig. 3) reveals that fundamentally different processes shape on-axis (red bars) and off-axis (blue bars) flow. On-axis, most upflow occurs between 300 °C and 450 °C, whereas downflow is strongest between 100 °C and 200 °C. These preferred temperature ranges are controlled by the thermodynamic properties of water<sup>4,24</sup>. Upflow occurs at temperatures around 425 °C, at which (for the prevailing pressures) the ability of water to transport energy by buoyancy-driven flow (that is, the fluid’s ‘fluxibility’<sup>24</sup>) is maximum. This also explains why even hotter fluids remain at depth and do not directly feed the on-axis plumes (see Figs 1d and 2c). Extending the fluxibility concept to predict the most efficient recharge flow<sup>4</sup> yields temperatures between 100 °C and 200 °C, which is where we also identify the strongest on-axis recharge flow (Fig. 2b). The energy required to heat the recharge flow is provided by the rising plumes, explaining why recharge is strongest adjacent to hot plumes (see Fig. 2c). The mass flux distribution for the off-axis flow component (blue bars in Fig. 2b) differs significantly,



**Figure 2 | Mass flux analysis.** **a**, Vertical mass flux of discharging and recharging fluid flow, evaluated 300 m below the sea floor, as a function of distance to the ridge axis. All fluid discharge, but only ~60% of recharge, occurs within a distance of 1.1 km of the ridge axis. The remaining ~40% recharge comes from fluids infiltrating the crust at greater distances. **b**, Vertical mass flux as a function of fluid temperature for on-axis (red) and off-axis (blue) flow components. Regions within the highly permeable pillow basalt layer and near

the top of the melt lens have been excluded for clarification (see Extended Data Fig. 3). Percentage values are with respect to total upflow and downflow within the entire domain, respectively. **c**, Vertical fluid mass flux shows on-axis thermal plumes at temperatures below ~450 °C that are surrounded by strong recharge flow (dark blue) at temperatures above 100 °C—both effects are consequences of pressure- and temperature-dependent fluid properties<sup>4,24</sup>.

with upward mass flux extending to much higher temperatures and most of the recharge flow occurring at colder temperatures. This implies that factors in addition to the fluid's fluxibility also control the deeper flow. Almost half of off-axis recharge flow occurs below 50 °C simply because no off-axis thermal plumes are present to heat the recharge. The extensive area of pervasive off-axis recharge compensates for these thermodynamically inefficient cold conditions. Upflow within the lower crust occurs over a broader range of higher temperatures and decreases towards 700 °C because permeability is reduced at the brittle–ductile transition (Fig. 2b, Extended Data Fig. 1). Hence, the off-axis upflow along the flanks of the hot slot is predominantly shaped by the crustal permeability structure, whereas the on-axis circulation forms thermal plumes more strongly controlled by the thermodynamic properties of water.

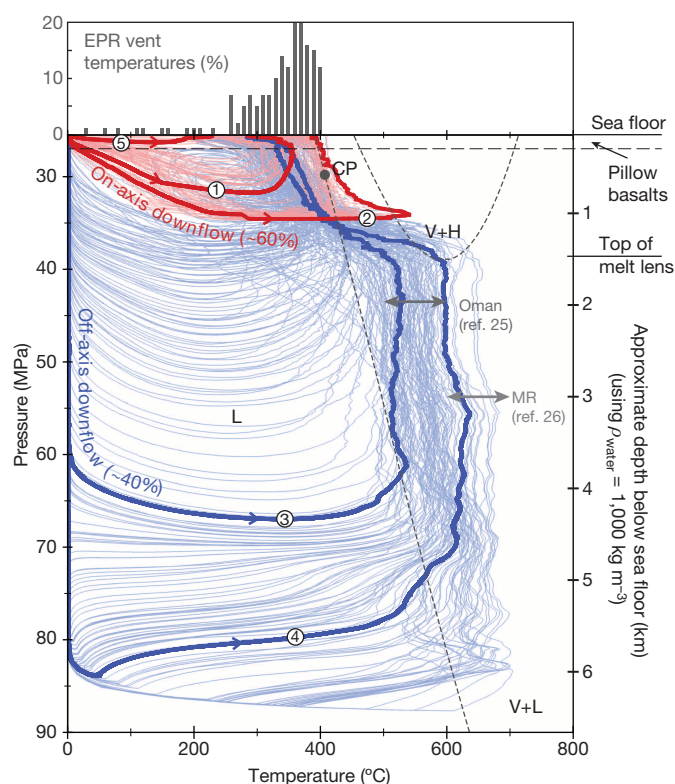
Pressure–temperature paths of numerical fluid tracers (Fig. 3) further highlight the characteristics of on-axis (red paths) and off-axis (blue paths) fluid circulation. With fluxibility limiting the temperatures in rising plumes above the melt lens to values below about 450 °C, the hot (up to 700 °C) deep-rooted fluids migrating towards the top of the melt lens cannot rise further until they cool below about 450 °C by mixing with the on-axis component. This may explain why the maximum pressures of last fluid–rock equilibration correlate very well with the depth of the axial melt lens<sup>16</sup>, even though a significant mass fraction of the venting fluids may actually have migrated through much deeper and hotter parts of the oceanic crust before venting. The pressure–temperature paths of on- and off-axis components also resolve an apparent discrepancy between *in situ* measurements of vent temperatures at the EPR and temperatures of hydrothermal circulation estimated from fluid inclusions in gabbros from the Oman ophiolite<sup>25</sup> and the extinct Mathematically ridge<sup>26</sup> (Fig. 3). Despite the clear upper limit in observed axial vent temperatures<sup>27</sup>, fluid inclusion data show much higher entrapment temperatures that are strikingly consistent with the temperatures our 3D model predicts for most deep off-axis upflow.

Our simulations have a total hydrothermal energy flux of about  $9 \times 10^7$  W per kilometre ridge axis, which is in excellent agreement with estimates for the total magmatic energy input<sup>28</sup> at a ridge with a full spreading rate of 11 cm yr<sup>-1</sup>. The model allows us to quantify the energy fluxes from upper and lower crust into the ocean. Despite contributing only 40% to the total fluid mass flux, the 500 °C–700 °C upflow in the lower crust transports about 70% of the total energy released at the

hydrothermal vent sites because it carries more thermal energy per kilogram of fluid than the <450 °C hot thermal plumes above the melt lens. This implies that not only is heat efficiently mined from the lower crust, but also that deep off-axis circulation may be the dominant process that cools the oceanic crust. Furthermore, the relative contributions of on- and off-axis hydrothermal flow to the total energy flux from crust to ocean are linked to the depth of the axial melt lens, where the transition in flow structure from hot (mainly permeability-controlled) flow in the lower gabbros to cooler (mainly fluid-property-controlled) flow in plumes occurs.

This new hybrid hydrothermal flow structure reconciles previously incompatible observations and models that suggested either strong on-axis<sup>4,5</sup> or deeper ridge-perpendicular hydrothermal circulation<sup>6,10,15</sup>. We find that both modes exist and naturally merge into a single hybrid flow structure in our whole-crust numerical experiments. The on-axis flow component is more vigorous and contributes slightly more to the total mass flux. It controls the depth of the melt lens and explains why high-temperature vent systems associated with fast-spreading ridges have so far been observed only directly on-axis. The deeper off-axis flow component provides about 70% of the hydrothermal energy flux and is what makes the overall thermal structure of the young oceanic crust consistent with seismic tomography data<sup>10</sup> and the pervasive crustal-scale hydrothermal convection inferred from *in situ* analysis of the Oman ophiolite<sup>7,15</sup>. This implies that the entire oceanic crust experiences extensive high-temperature hydrothermal circulation, providing a mechanism to scavenge elements from the lower crust. Recent mass balance calculations reveal that the amounts of metal potentially mobilized by hydrothermal flow at ridges<sup>12</sup> are far greater than estimates based on the size of known ridge-linked massive sulphide deposits<sup>29</sup>. Our simulations predict a cooling of about 200 °C of deeper off-axis fluids upon mixing with the on-axis component above the melt lens, which would lead to strong precipitation of dissolved metals near the roots of the sheeted dykes. Entrainment of cold sea water is likely to result in further re-deposition of metal within the pillow basalt layer. Finally, chemical differences between individual vents of the same vent field as observed at the neighbouring P and Bio9 vents<sup>30</sup> at EPR 9° 50' N could result from mixing processes between fluids that obtained distinct chemical signatures while cycling through the conjoined limbs of a single hybrid hydrothermal system.





**Figure 3 | Comparison between observed and modelled hydrothermal fluid temperatures.** *In situ* measurements of exit temperatures at EPR vents (see Methods) show a preferred range of 300 °C–400 °C. Temperature estimates based on fluid inclusions in gabbros from the Oman ophiolite<sup>25</sup> and the extinct Mathematician ridge<sup>26</sup> (MR) are significantly higher. The pressure–temperature paths of on-axis (red) and off-axis (blue) fluid tracers in our 3D model explain the discrepancy. Highlighted paths 1 and 2 show warm recharge flow before ascending in plumes at temperatures below ~450 °C. In contrast, off-axis recharge flow (paths 3 and 4) is cold until reaching the steep thermal gradients at depth. Most of the subsequent upward flow occurs at 500 °C–700 °C—the temperature range suggested by fluid inclusion data. Fluid properties at these high temperatures are thermodynamically inefficient for plume formation<sup>24</sup> and off-axis fluids must cool by mixing with the on-axis component before they are able to feed a plume. Path 5 shows entrainment of cold sea water into a thermal plume within the pillow basalt layer. Dashed lines mark phase regions of sea water for orientation (L, liquid; V, vapour; H, halite) and CP marks the critical point of sea water.

## METHODS SUMMARY

We constructed a 2D numerical model<sup>18</sup> representing a vertical cross-section perpendicular to a fast-spreading ridge to solve for the long-term thermal evolution of oceanic crust and mantle. Hydrothermal convection of pure water is coupled to crustal accretion by the temperature field. Fluid flow is treated as Darcy flow through a porous medium with parameterized permeability, and we solve for temperature, pore pressure and fluid velocity using an implicit finite-element method. Deformation of the crust is formulated as viscous creep of an incompressible medium and a dilation term is used to simulate crustal accretion in the melt lens and the dyking region. 2D simulations ran until a quasi-steady state was reached, typically after 200,000–500,000 years. The purpose of this model is to define a permeability field that leads to a thermal structure matching the one inferred from seismic data<sup>10</sup>. Furthermore, the melt lens (corresponding to the 1,200 °C isotherm) should be located at the observed depth of 1,400 m. The obtained permeability field forms the basis for the 3D model, which covers the crustal volume above the 1,000 °C isotherm in the 2D results and extends 6 km along the ridge and 8 km to each side of the ridge. In- and outflow are allowed through the top boundary only. The numerical formulation of hydrothermal flow is very similar to that used in the 2D model, but solving the much larger system of equations required parallelization of the 3D model and iterative solution algorithms<sup>31</sup>. Each 3D calculation starts at cold crustal conditions, with the exception of a hot boundary layer along the bottom boundary that forms after 300 years of purely conductive heating before model initiation. Each 3D calculation simulated 2,000–8,000 years, after which vent field locations stabilized and discharging mass fluxes became nearly constant.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 December 2013; accepted 11 February 2014.

- Stein, C. A. & Stein, S. Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *J. Geophys. Res.* **99**, 3081–3095 (1994).
- Baker, E., Chen, Y. & Phipps Morgan, J. The relationship between near-axis hydrothermal cooling and the spreading rate of mid-ocean ridges. *Earth Planet. Sci. Lett.* **142**, 137–145 (1996).
- Singhal, B. B. S. & Gupta, R. P. *Applied Hydrogeology of Fractured Rocks* (Kluwer Academic, 1999).
- Coumou, D., Driesner, T. & Heinrich, C. A. The structure and dynamics of mid-ocean ridge hydrothermal systems. *Science* **321**, 1825–1828 (2008).
- Tolstoy, M., Waldhauser, F., Bohnenstiehl, D. R., Weekly, R. T. & Kim, W. Y. Seismic identification of along-axis hydrothermal flow on the East Pacific Rise. *Nature* **451**, 181–184 (2008).
- Cathles, L. M. A capless 350 °C flow zone model to explain megaplumes, salinity variations, and high-temperature veins in ridge axis hydrothermal systems. *Econ. Geol.* **88**, 1977–1988 (1993).
- Nicolas, A., Mainprice, D. & Boudier, F. High-temperature seawater circulation throughout crust of oceanic ridges: a model derived from the Oman ophiolites. *J. Geophys. Res.* **108** (B8), 2371, <http://dx.doi.org/10.1029/2002JB002094> (2003).
- Johnson, H. P., Becker, K. & Von Herzen, R. Near-axis heat flow measurements on the northern Juan De Fuca Ridge: implications for fluid circulation in oceanic crust. *Geophys. Res. Lett.* **20**, 1875–1878 (1993).
- Detrick, R. S. et al. Multi-channel seismic imaging of a crustal magma chamber along the East Pacific Rise. *Nature* **326**, 35–41 (1987).
- Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at 9° 30' N. *J. Geophys. Res.* **105**, 23537–23555 (2000).
- Van Tongeren, J. A., Kelemen, P. B. & Hanghøj, K. Cooling rates in the lower crust of the Oman ophiolite: Ca in olivine, revisited. *Earth Planet. Sci. Lett.* **267**, 69–82 (2008).
- Cathles, L. M. What processes at mid-ocean ridges tell us about volcanogenic massive sulfide deposits. *Mineral. Dep.* **46**, 639–657 (2011).
- Cherkaoui, A. S. M., Wilcock, W. S. D., Dunn, R. A. & Toomey, D. R. A numerical model of hydrothermal cooling and crustal accretion at a fast spreading mid-ocean ridge. *Geochem. Geophys. Geosyst.* **4**, 8616, <http://dx.doi.org/10.1029/2001GC000215> (2003).
- Fisher, A. T. in *Dahlem Workshop Report: Energy and Mass Transfer in Marine Hydrothermal Systems* (eds Halbach, P. E., Tunncliffe, V. & Hein, J. R.) 29–52 (Dahlem Univ. Press, 2003).
- Bosch, D. et al. Deep and high-temperature hydrothermal circulation in the Oman ophiolite—petrological and isotopic evidence. *J. Petrol.* **45**, 1181–1208 (2004).
- Fontaine, F. J., Wilcock, W. S. D., Foustoukos, D. E. & Butterfield, D. A. Si-Cl geothermobarometer for the reaction zone of high-temperature, basaltic-hosted mid-ocean ridge hydrothermal systems. *Geochem. Geophys. Geosyst.* **10**, Q05009 (2009).
- Coumou, D., Driesner, T., Geiger, S. & Paluszny, A. High-resolution three-dimensional simulations of mid-ocean ridge hydrothermal systems. *J. Geophys. Res.* **114**, B07104 (2009).
- Theissen-Krah, S., Iyer, K., Rüpke, L. H. & Morgan, J. P. Coupled mechanical and hydrothermal modeling of crustal accretion at intermediate to fast spreading ridges. *Earth Planet. Sci. Lett.* **311**, 275–286 (2011).
- Lister, C. R. B. On the penetration of water into hot rock. *Geophys. J. R. Astron. Soc.* **39**, 465–509 (1974).
- Boudier, F., Nicolas, A. & Mainprice, D. Does anisotropy of thermal contraction control hydrothermal circulation at the Moho level below fast spreading oceanic ridges? *Int. Geol. Rev.* **47**, 101–112 (2005).
- Carlson, R. L. The effect of hydrothermal alteration on the seismic structure of the upper oceanic crust: evidence from Holes 504B and 1256D. *Geochem. Geophys. Geosyst.* **12**, Q09013 (2011).
- Hirth, G., Escartin, J. & Lin, J. in *Faulting and Magmatism at Mid-Ocean Ridges* (eds Buck, W. R., Delaney, P. T., Karson, J. A. & Lagabrielle, Y.) **106**, 291–303 (American Geophysical Union, 1998).
- Harding, A. J., Kent, G. M. & Orcutt, J. A. A multichannel seismic investigation of upper crustal structure at 9° N on the East Pacific Rise: implications for crustal accretion. *J. Geophys. Res.* **98**, 13925–13944 (1993).
- Jupp, T. & Schultz, A. A thermodynamic explanation for black smoker temperatures. *Nature* **403**, 880–883 (2000).
- Nehlig, P. Interactions between magma chambers and hydrothermal systems: oceanic and ophiolitic constraints. *J. Geophys. Res.* **98**, 19621–19633 (1993).
- Vanko, D. A. Temperature, pressure, and composition of hydrothermal fluids, with their bearing on the magnitude of tectonic uplift at mid-ocean ridges, inferred from fluid inclusions in oceanic layer 3 rocks. *J. Geophys. Res.* **93**, 4595–4611 (1988).
- Hannington, M. D., de Ronde, C. E. J. & Petersen, S. Sea-floor tectonics and submarine hydrothermal systems. *Econ. Geol.* **100th Anniversary Volume**, 111–141 (2005).
- Sinha, M. C. & Evans, R. L. in *Mid-Ocean Ridges: Hydrothermal Interactions Between the Lithosphere and Oceans* (eds German, C., Lin, J. & Parson, L.) **148**, 19–62 (American Geophysical Union, 2004).

29. Hannington, M., Jamieson, J., Monecke, T., Petersen, S. & Beaulieu, S. The abundance of seafloor massive sulfide deposits. *Geology* **39**, 1155–1158 (2011).
30. Fornari, D. *et al.* The East Pacific Rise between 9°N and 10°N: twenty-five years of integrated, multidisciplinary oceanic spreading center studies. *Oceanography* **25**, 18–43 (2012).
31. Hasenclever, J. *Modeling Mantle Flow and Melting Processes at Mid-Ocean Ridges and Subduction Zones — Development and Application of Numerical Models*. PhD thesis, Hamburg Univ. (2010).

**Acknowledgements** We thank reviewers T. Driesner and P. Johnson for comments that led to more insights into hydrothermal energy transport and improved the manuscript.

**Author Contributions** J.H. and J.P.M. developed the 3D numerical model. J.H. carried out the 3D simulations, did the post-processing and designed the figures. 2D simulations were done by S.T.-K. and J.H. (using the 2D model developed by S.T.-K., L.H.R., K.I. and J.P.M.). J.H. and L.H.R. wrote the initial manuscript, to which J.P.M., S.T.-K., K.I., S.P. and C.W.D. contributed geological and thermodynamic implications. Figures and text were edited and improved by all authors. All authors discussed the results and implications at all stages of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H. ([jhasenclever@geomar.de](mailto:jhasenclever@geomar.de)).

# Plate tectonics, damage and inheritance

David Bercovici<sup>1</sup> & Yanick Ricard<sup>2</sup>

**The initiation of plate tectonics on Earth is a critical event in our planet's history. The time lag between the first proto-subduction (about 4 billion years ago) and global tectonics (approximately 3 billion years ago) suggests that plates and plate boundaries became widespread over a period of 1 billion years. The reason for this time lag is unknown but fundamental to understanding the origin of plate tectonics. Here we suggest that when sufficient lithospheric damage (which promotes shear localization and long-lived weak zones) combines with transient mantle flow and migrating proto-subduction, it leads to the accumulation of weak plate boundaries and eventually to fully formed tectonic plates driven by subduction alone. We simulate this process using a grain evolution and damage mechanism with a composite rheology (which is compatible with field and laboratory observations of polycrystalline rocks<sup>1,2</sup>), coupled to an idealized model of pressure-driven lithospheric flow in which a low-pressure zone is equivalent to the suction of convective downwellings. In the simplest case, for Earth-like conditions, a few successive rotations of the driving pressure field yield relic damaged weak zones that are inherited by the lithospheric flow to form a nearly perfect plate, with passive spreading and strike-slip margins that persist and localize further, even though flow is driven only by subduction. But for hotter surface conditions, such as those on Venus, accumulation and inheritance of damage is negligible; hence only subduction zones survive and plate tectonics does not spread, which corresponds to observations. After plates have developed, continued changes in driving forces, combined with inherited damage and weak zones, promote increased tectonic complexity, such as oblique subduction, strike-slip boundaries that are subparallel to plate motion, and spalling of minor plates.**

The emergence of plate tectonics is arguably Earth's defining moment. How our planet, alone amongst known terrestrial bodies, evolved the unique plate-tectonic form of mantle convection remains enigmatic<sup>3</sup>. Although geochemical and petrological evidence for tectonic activity in Earth's deepest past is difficult to interpret<sup>4</sup>, analysis of zircons implies the existence of continental crust and formation of granites more than 4 billion years (4 Gyr) ago<sup>5,6</sup>, possibly related to subduction zone melting. Similar evidence suggests that subduction-related crustal production existed 3.6 Gyr ago or earlier<sup>7,8</sup>. However, geochemical and petrologic observations imply that tectonics was not widespread until 3.0–2.7 Gyr ago<sup>9,10</sup>. Whether plate tectonics spread globally over the intervening 1-Gyr interval is of fundamental importance in understanding the emergence of plate tectonics. Moreover, whether the continued accumulation of plate boundaries increased tectonic complexity through geological time remains an open question. Here we propose a plate generation model to argue that transient proto-subduction and accumulated lithospheric damage allowed for the unique emergence and evolution of plate tectonics on Earth.

Plate generation theory generally treats the formation of weak plate boundaries and strong plates by invoking self-weakening feedback mechanisms<sup>3</sup>. Models prescribing steady-state weakening rheologies (for example, viscoplasticity)<sup>11–13</sup> often produce instantaneous plate-like flow, but are not necessarily consistent with rock-deformation experiments, and do not contain the physics to allow dormancy and accumulation of weak zones that become re-activated to form new plate boundaries<sup>14,15</sup>.

Damage models of plate generation have, however, been developed to treat self-weakening as well as the evolution of weakness, while also being compatible with laboratory and field observations<sup>1</sup>.

Damage mechanics is a continuum theory for how deformation affects material strength, usually by the formation of micro-cracks or reduction of grain size, either of which weakens the material<sup>16</sup>. The pervasiveness of mantle-lithosphere peridotitic mylonites—where localized deformation correlates with reduced mineral grain sizes—at tectonic margins<sup>17,18</sup> indicates that a self-weakening feedback between grain reduction by damage and grain-size-dependent strength (for example, diffusion creep)<sup>19,20</sup> at mid-lithospheric depths is key to plate-boundary formation. This grain-damage feedback mechanism is most plausible and evident in multi-phase or polycrystalline lithospheric rocks (for example, peridotite, which is a mixture of mostly olivine and pyroxene): grain growth in such rocks—which strengthens and heals the material—is known from laboratory and field observations to be impeded by pinning (that is, blocking of grain-boundary migration) by the interface between phases<sup>17,21,22</sup>. Thus, our lithospheric damage mechanism is a coupled evolution model that describes the competition between damage and healing for both grains and interfaces in two-phase assemblages, and where interface pinning both limits grain growth and healing, and promotes damage and shear-localization<sup>1,2</sup> (see Methods for further details).

Our model medium is composed of two components or phases with comparable properties, combined into a 60–40% mixture (as with olivine–pyroxene ratios in peridotite), wherein the grain sizes and the roughness of the interface between phases change according to our evolution model<sup>1,2</sup>. The two-phase grained continuum is incorporated into a horizontal viscous thin-sheet model of convective-type flow in the lithosphere, assuming its strength is dominated by highly viscous flow and mylonitic behaviour at a mid-lithosphere depth (this assumption, however, underestimates weakening and localization at divergent zones due to lithospheric thinning and melting). Flow velocity  $v$  is compelled by imposed sublithospheric pressure anomalies  $P$ ; high or low pressures are akin to mantle convective currents impinging on or separating from this layer. The material rheology is a composite of dislocation creep (that is, viscosity goes inversely as stress squared), which dominates deformation for large grains, and diffusion creep (that is, viscosity goes as grain-size cubed), which dominates for small grains. As deformation and damage drive grain reduction<sup>1</sup>, diffusion creep eventually dominates, causing the mixture viscosity  $\bar{\mu}$  to reduce where deformation is large, which thus focuses deformation and leads to a self-weakening feedback. If  $\bar{\mu}$  were constant, the pressure would drive only divergence or convergence (over positive or negative  $P$ , respectively), or sources and sinks of mass in the thin layer, akin to ridges and subduction zones; the divergence rate is denoted by  $S = \bar{\nabla} \cdot v$  where  $\bar{\nabla}$  is the horizontal two-dimensional gradient operator (see Methods). But with the nonlinear rheology of our medium, horizontal shearing and strike-slip motion can also arise; the vorticity  $\Omega = \hat{z} \cdot \nabla \times v$  (where  $\hat{z}$  is the vertical unit vector) represents shearing because it measures velocity gradients perpendicular to the flow direction. Moreover, the driving pressure  $P$  can also interact with the nonlinear rheology to excite passive divergence that is not directly driven by any pressure anomaly. Focused divergence  $S$  and vorticity  $\Omega$  are important metrics of plate-like flow because

<sup>1</sup>Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520-8109, USA. <sup>2</sup>Laboratoire de Géologie de Lyon, Université de Lyon 1, CNRS, ENS-Lyon, 69622 Villeurbanne Cedex, France.



they represent convergent, divergent and strike-slip motion on plate boundaries<sup>3</sup>. See Methods for further details.

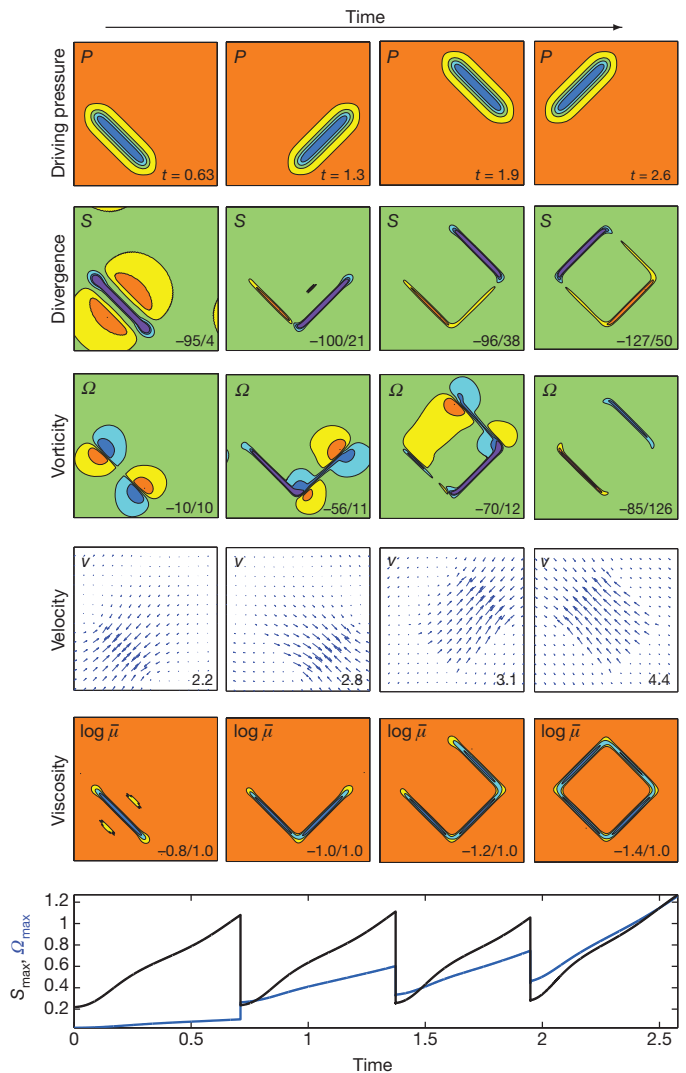
The model input requires several dimensionless ratios or numbers, which are derived from known physical quantities and laboratory experiments (see Methods). The key parameters controlling the efficiency of plate generation are the damage and healing numbers for both grains and interface; these numbers represent the rates of comminution and coarsening (for either grains or interface), respectively, relative to the rate of deformation. There are three damage numbers, that is, for the interface between phases and the grains of each phase, but they are typically equal to each other, and thus we denote them by one number  $D$ ; for Earth's conditions,  $D \approx 100$ . There are likewise three healing numbers,  $C_i$  for grain growth (where  $i = 1$  or  $2$  depending on the phase) and  $C_l$  for interface coarsening; unimpeded grain growth is typically much faster than interface healing<sup>1</sup> and for Earth's conditions we infer that both phases have  $C_i \approx 1$ , and the interface has  $C_l \approx 10^{-5}$ . Several tests and comparisons of basic rheological mechanisms are discussed in Methods and shown in Extended Data.

We first consider an idealized model of an early Archaean, single proto-subduction zone or downwelling, by imposing a pressure field  $P$  with one focused low-pressure region evolving with time (Fig. 1 top row), which is representative of time-dependent intermittent downwelling inferred to occur in the Archaean<sup>23</sup>. In particular, the first downwelling is akin to a lithospheric drip, which typically cannot readily draw in lithospheric mass, and is therefore ephemeral. A new similar lithospheric downwelling is prescribed to nucleate perpendicularly off a pre-weakened end of the first downwelling, because other orientations would probably be swept away by the initial flow field. The new downwelling is also relatively short-lived, and a third and then fourth downwellings similarly nucleate in turn.

Even with a damage rheology, the first low-pressure region leads only to a convergence zone or sink of lithospheric mass (Fig. 1 left column). During the next downwelling stage, the new flow is influenced by the damaged weak zone left by the original downwelling, whose signature survives because of slow healing; this weak zone becomes a locus of vorticity and strike-slip motion (Fig. 1, second column). In the third stage, the weakness left from the second downwelling is adopted as a strike-slip margin, while that from the original downwelling becomes a focused spreading centre (Fig. 1, third column). As the process is repeated one last time, the accumulated weak zones comprise narrow divergent and strike-slip boundaries, which complement the driving subduction zone to make one contiguous weak plate margin bounding a rapidly moving tectonic plate (Fig. 1, forth column). Moreover, with each subduction zone jump, the divergence field  $S$  rebuilds to approximately the same magnitude, but the vorticity  $\Omega$  progressively grows and finally has a magnitude comparable to that of  $S$  (Fig. 1, bottom row). Although the calculation ceases after the last shift in the downwelling, we speculate that the final plate-like flow would organize convection and limit further evolution of cold downwellings and drips by sweeping them into the existing subduction zone.

Although this case is highly idealized, it shows that a fully developed plate can evolve from a downwelling only—where none of the boundaries other than the subduction zone are actively driven—given transience and re-nucleation of the downwelling, and a damage mechanism that provides localization and longevity of weak zones. With these basic ingredients, single proto-subduction zones could build and propagate plate tectonics by accumulating weak zones to be inherited as passive plate boundaries. The time to localize weak zones is several tens of Myr, thus if the time to form new drip downwellings is, with damage<sup>24</sup>, a few hundred Myr, then there is sufficient time to develop a full plate in the 1 Gyr between initial proto-subduction and global plate tectonics.

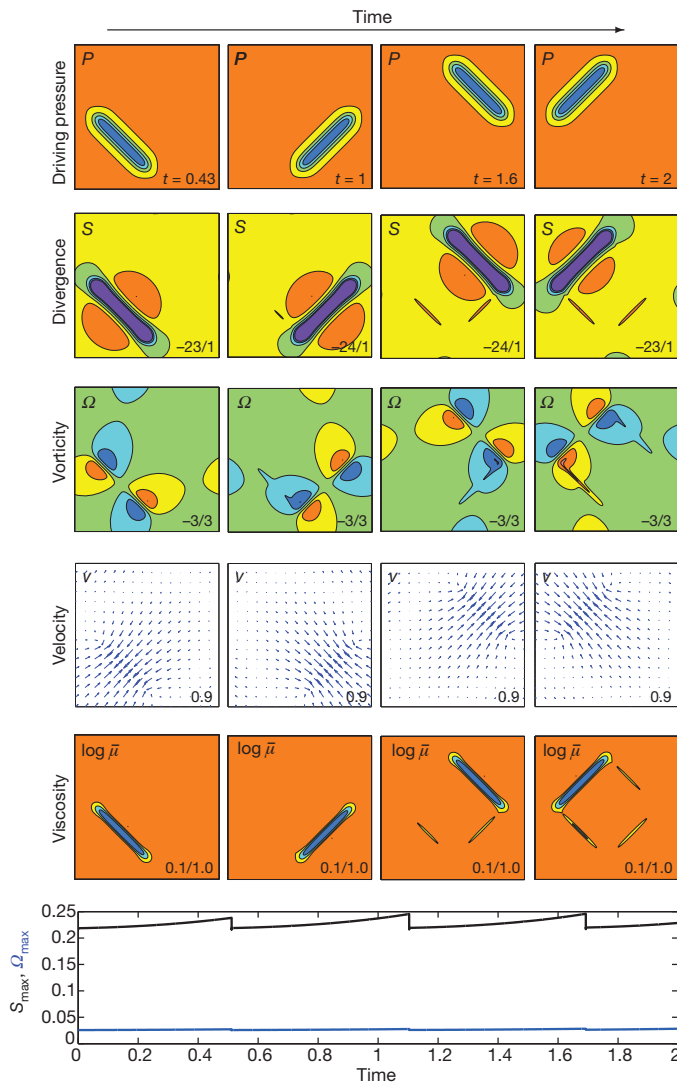
A similar exercise using hotter Venusian surface conditions permits a comparison to Earth's putative twin<sup>25–27</sup>. Using the same material properties but a lithosphere temperature elevated by 200–400 K above the Earth-like case, the damage number  $D$  is reduced by a factor of about 10, and the healing numbers  $C_i$  and  $C_l$  are increased by a factor of up to



**Figure 1 | Lithospheric flow model with damage driven by intermittent proto-subduction.** The low in the two-dimensional horizontal pressure field  $P$  represents a convective downwelling, rotations of which (top row, with time  $t$  indicated) model nucleation of new proto-subduction. Weak regions inherited from previous downwellings localize into passive plate margins, indicated by divergence  $S$ , strike-slip vorticity  $\Omega$ , horizontal velocity  $v$  and viscosity  $\bar{\mu}$  (rows 2–4). The final state is a four-sided plate, driven by subduction only. Bottom row shows evolution of maximum divergence and vorticity. Scales of distance, time and pressure are  $\sim 5,000$  km,  $\sim 50$  Myr and  $\sim 300$  MPa. Extrema (multiplied by 100, except for  $\log \bar{\mu}$ ) are indicated on each frame, save  $P$ , which is always 0–1. See Extended Data Fig. 1 for contour details.

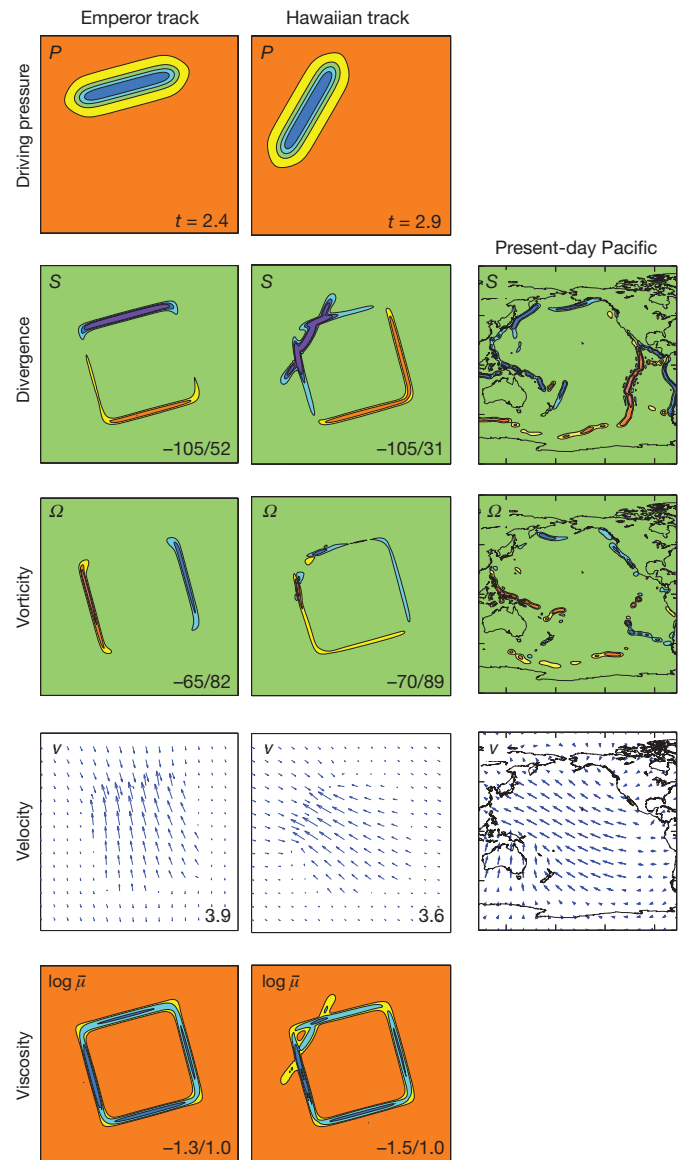
10 (see Methods). In this case, as a downwelling migrates to various positions (see Fig. 2), only very faint weak zones accumulate because damage itself is weaker while healing is stronger, thereby resulting in zones of passive divergence and strike-slip vorticity about an order of magnitude weaker than the convergence rate. The final flow field is dominated by convergence, giving the appearance of a subduction-only surface. This result provides a simple explanation for why Venus possibly has subduction zones<sup>28</sup>, but no extant plate tectonics.

After the establishment of global plate tectonics and continental growth on Earth, approximately 3 Gyr ago, plate reorganizations became more complicated. Nevertheless, we explore an idealized model of present-day plate motion changes and the resulting accumulation of complexity by inherited weak zones. The experiment is an abstraction of the change in motion of the Pacific plate associated with the 47-Myr-ago Emperor-Hawaiian bend<sup>2,29</sup>. The initial condition has one low-pressure or 'subduction' zone driving a plate-like flow field, similar to the final



**Figure 2 | Lithospheric flow model for Venus.** Quantities shown are the same as in Fig. 1 but with parameters appropriate for Venus' higher surface temperature, leading to damage parameter  $\mathcal{D} = 10$  (instead of 100 for the Earth case) and healing numbers  $C_i = 10$  and  $C_l = 10^{-4}$  (instead of  $C_i = 1$  and  $C_l = 10^{-5}$  for the Earth case). Length scale is the same as in Fig. 1, the time and pressure scales are reduced by a factor of about 10 (assuming a 200 K increase in mid-lithosphere temperature from the Earth model<sup>23,27</sup>).

stage shown in Fig. 1. This pressure field initially drives motion north with azimuth  $15^\circ$  west of north (W15N), which is approximately the trajectory of the Emperor seamount chain (Fig. 3 first column). The pressure gradient is then rotated to W60N, roughly the azimuth of the Hawaiian chain (Fig. 3 second column). After the rotation in  $P$ , the viscosity  $\bar{\mu}$  and thus both the divergence  $S$  and vorticity  $\Omega$  retain 'memory' of the prior plate configuration. Although highly idealized, the experiment yields several features suggestive of present-day Pacific plate motion (Fig. 3 third column). First, the passive divergent zones inherit the older plate geometry and form two ridges wrapping around the southeast corner of the plate, similar to the East Pacific Rise and the Pacific-Antarctic Ridge. The new vorticity field inherits all four of the pre-rotation weak zones, and continues to grow in magnitude; the resulting strike-slip motion is non-orthogonal to the new subduction zone, leading to oblique convergence along the northern boundary, similar to the Aleutian trench, strike-slip motion on the eastern boundary similar to the San Andreas system, and oblique spreading as on the southern Pacific plate. The northern junction between the old and new subduction zones is suggestive of the sharp Aleutian-Kurile arc corner,



**Figure 3 | Lithospheric flow model with Pacific-like rotation.** Quantities shown are the same as in Fig. 1 with the same Earth-like parameters. The initial configuration has established plate boundaries, as in the final step of Fig. 1, but in a W15N direction analogous to the Emperor seamount chain trajectory (left column). The pressure field then rotates  $45^\circ$  to W60N analogous to the present-day Hawaiian island chain direction (middle column). Divergence, vorticity and velocity of the Earth's present-day Pacific plate (computed up to spherical harmonic degree 64, using the Nuvel1 plate model<sup>30</sup>) are shown for comparison (right column).

and the truncated convergent corner of the prior plate is reminiscent of the Philippine plate (see Fig. 3 second column, viscosity field at bottom). In the end, the inheritance of damaged zones by lithospheric flow after a simple change in the slab-pull force leads to several basic features suggestive of Earth's actual Pacific plate.

In summary, we posit that the emergence and evolution of plate tectonics arise by the interaction of a realistic polycrystalline grain-damage mechanism<sup>1,2</sup> and time-dependent subduction, which together allow plate boundary localization and longevity.

## METHODS SUMMARY

The theoretical model involves flow in a thin layer of grained, viscous two-phase fluid, wherein the phases represent olivine and pyroxene in a peridotite lithosphere. The layer's rheology is a composite between dislocation and diffusion creep, depending on whether grains are, respectively, bigger or smaller than a critical size. Grain

evolution is coupled to the evolution of the interface between the two phases, which imposes Zener pinning surfaces that block normal grain growth depending on the sharpness of these surfaces. The interface sharpness evolves by the competition between surface-tension-driven coarsening and damage-driven sharpening (that is, distortion, mixing and rending), and sharpened pinning surfaces further block grain growth and facilitate grain-size reduction. When grains shrink enough, they undergo diffusion creep, which depends strongly on grain size, thus leading to a self-softening feedback; that is, deformation reduces grain size and viscosity, which enhances deformation and damage, which reduces grain size and so on. Pinning also prevents healing of small-grained weak zones, depending on temperature; thus weak zones can be inherited and reactivated even after falling dormant. Flow in the thin layer is driven by a low-pressure zone imposed at its base, which represents the pull of a convective downwelling. In a pristine lithosphere, only convergent flow is driven into the low-pressure 'downwelling' and generates a damaged weak zone there. The pressure field is then migrated, and the previously formed, slowly-healing damaged weak zones are inherited by the lithospheric flow and reactivated to form focused divergent or strike-slip zones, or combinations thereof, depending on their alignment. Thus, focused plate boundaries of all types arise by coupling time-dependent downwelling with polycrystalline grain damage. Moreover, as the healing rate of weak zones depends on surface temperature, colder surfaces, as on Earth, promote more intense weak zone formation and inheritance than do much hotter surfaces, as on Venus.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 September 2013; accepted 21 January 2014.

Published online 6 April 2014.

- Bercovici, D. & Ricard, Y. Mechanisms for the generation of plate tectonics by two-phase grain-damage and pinning. *Phys. Earth Planet. Inter.* **202–203**, 27–55 (2012).
- Bercovici, D. & Ricard, Y. Generation of plate tectonics with two-phase grain-damage and pinning: source-sink model and toroidal flow. *Earth Planet. Sci. Lett.* **365**, 275–288 (2013).
- Bercovici, D. The generation of plate tectonics from mantle convection. *Earth Planet. Sci. Lett.* **205**, 107–121 (2003).
- Korenaga, J. Initiation and evolution of plate tectonics on Earth: theories and observations. *Annu. Rev. Earth Planet. Sci.* **41**, 117–151 (2013).
- Harrison, T. M. et al. Heterogeneous Hadean hafnium: evidence of continental crust at 4.4 to 4.5 Ga. *Science* **310**, 1947–1950 (2005).
- Shirey, S., Kamber, B., Whitehouse, M., Mueller, P. & Basu, A. in *When Did Plate Tectonics Begin on Planet Earth?* (eds Condie, K. & Pease, V.) 1–29 (Special Paper 440, Geological Society of America, 2008).
- Hopkins, M. D., Harrison, T. M. & Manning, C. E. Constraints on Hadean geodynamics from mineral inclusions in >4Ga zircons. *Earth Planet. Sci. Lett.* **298**, 367–376 (2010).
- Polat, A., Appel, P. W. & Fryer, B. J. An overview of the geochemistry of Eoarchean to Mesoarchean ultramafic to mafic volcanic rocks, SW Greenland: implications for mantle depletion and petrogenetic processes at subduction zones in the early Earth. *Gondwana Res.* **20**, 255–283 (2011).
- Condie, K. & Kröner, A. in *When Did Plate Tectonics Begin on Planet Earth?* (eds Condie, K. & Pease, V.) 281–294 (Special Paper 440, Geological Society of America, 2008).
- Shirey, S. B. & Richardson, S. H. Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* **333**, 434–436 (2011).
- Trompert, R. & Hansen, U. Mantle convection simulations with rheologies that generate plate-like behaviour. *Nature* **395**, 686–689 (1998).
- van Heck, H. & Tackley, P. Planforms of self-consistently generated plates in 3D spherical geometry. *Geophys. Res. Lett.* **35**, L19312 (2008).
- Foley, B. & Becker, T. Generation of plate-like behavior and mantle heterogeneity from a spherical, visco-plastic convection model. *Geochem. Geophys. Geosyst.* **10**, Q08001 (2009).
- Zhong, S., Gurnis, M. & Moresi, L. Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models. *J. Geophys. Res.* **103**, 15255–15268 (1998).
- Gurnis, M., Zhong, S. & Toth, J. in *History and Dynamics of Global Plate Motions* (eds Richards, M. A., Gordon, R. & van der Hilst, R.) 73–94 (Geophys. Monogr. Ser. Vol. 121, Am. Geophys. Union, 2000).
- Krajcinovic, D. *Damage Mechanics* (North-Holland, 1996).
- Warren, J. M. & Hirth, G. Grain size sensitive deformation mechanisms in naturally deformed peridotites. *Earth Planet. Sci. Lett.* **248**, 438–450 (2006).
- Skemer, P., Warren, J. M., Kelemen, P. B. & Hirth, G. Microstructural and rheological evolution of a mantle shear zone. *J. Petrol.* **51**, 43–53 (2010).
- Karato, S., Toriumi, M. & Fujii, T. Dynamic recrystallization of olivine single crystals during high temperature creep. *Geophys. Res. Lett.* **7**, 649–652 (1980).
- Austin, N. & Evans, B. Paleowattmeters: a scaling relation for dynamically recrystallized grain size. *Geology* **35**, 343–346 (2007).
- Hiraga, T., Tachibana, C., Ohashi, N. & Sano, S. Grain growth systematics for forsterite ± enstatite aggregates: effect of lithology on grain size in the upper mantle. *Earth Planet. Sci. Lett.* **291**, 10–20 (2010).
- Linckens, J., Herwegh, M., Müntener, O. & Mercolli, I. Evolution of a polyminerale mantle shear zone and the role of second phases in the localization of deformation. *J. Geophys. Res.* **116**, B06210 (2011).
- van Hunen, J. & Moya, J.-F. Archean subduction: fact or fiction? *Annu. Rev. Earth Planet. Sci.* **40**, 195–219 (2012).
- Paczowski, K., Bercovici, D., Landuyt, W. & Brandon, M. T. Drip instabilities of continental lithosphere: acceleration and entrainment by damage. *Geophys. J. Int.* **189**, 717–729 (2012).
- Lenardic, A., Jellinek, M. & Moresi, L.-N. A climate change induced transition in the tectonic style of a terrestrial planet. *Earth Planet. Sci. Lett.* **271**, 34–42 (2008).
- Landuyt, W. & Bercovici, D. Variations in planetary convection via the effect of climate on damage. *Earth Planet. Sci. Lett.* **277**, 29–37 (2009).
- Foley, B. J., Bercovici, D. & Landuyt, W. The conditions for plate tectonics on super-Earths: inferences from convection models with damage. *Earth Planet. Sci. Lett.* **331–332**, 281–290 (2012).
- Schubert, G. & Sandwell, D. A global survey of possible subduction sites on Venus. *Icarus* **117**, 173–196 (1995).
- Sharp, W. D. & Clague, D. A. 50-Ma initiation of Hawaiian-Emperor bend records major change in Pacific plate motion. *Science* **313**, 1281–1284 (2006).
- Argus, D. F. & Gordon, R. G. No-net-rotation model of current plate velocities incorporating plate motion model NUVEL-1. *Geophys. Res. Lett.* **18**, 2039–2042 (1991).

**Acknowledgements** D.B. acknowledges support from the National Science Foundation; Y.R. acknowledges support from the Agence Nationale de la Recherche. This work benefitted from discussions with S. Karato, G. Hirth, N. Coltice and B. J. Foley.

**Author Contributions** D.B. and Y.R. conceived the physical and mathematical model together. D.B. developed and deployed the computational model and was the lead author for the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.B. ([david.bercovici@yale.edu](mailto:david.bercovici@yale.edu)).



# Herbivores and nutrients control grassland plant diversity via light limitation

Elizabeth T. Borer<sup>1</sup>, Eric W. Seabloom<sup>1</sup>, Daniel S. Gruner<sup>2</sup>, W. Stanley Harpole<sup>3</sup>, Helmut Hillebrand<sup>4</sup>, Eric M. Lind<sup>1</sup>, Peter B. Adler<sup>5</sup>, Juan Alberti<sup>6</sup>, T. Michael Anderson<sup>7</sup>, Jonathan D. Bakker<sup>8</sup>, Lori Biederman<sup>3</sup>, Dana Blumenthal<sup>9</sup>, Cynthia S. Brown<sup>10</sup>, Lars A. Brudvig<sup>11</sup>, Yvonne M. Buckley<sup>12,13</sup>, Marc Cadotte<sup>14</sup>, Chengjin Chu<sup>15</sup>, Elsa E. Cleland<sup>16</sup>, Michael J. Crawley<sup>17</sup>, Pedro Daleo<sup>6</sup>, Ellen I. Damschen<sup>18</sup>, Kendi F. Davies<sup>19</sup>, Nicole M. DeCrappeo<sup>20</sup>, Guozhen Du<sup>15</sup>, Jennifer Firn<sup>21</sup>, Yann Hautier<sup>1</sup>, Robert W. Heckman<sup>22</sup>, Andy Hector<sup>23</sup>, Janneke HilleRisLambers<sup>24</sup>, Oscar Iribarne<sup>6</sup>, Julia A. Klein<sup>10</sup>, Johannes M. H. Knops<sup>25</sup>, Kimberly J. La Pierre<sup>26</sup>, Andrew D. B. Leakey<sup>27</sup>, Wei Li<sup>3</sup>, Andrew S. MacDougall<sup>28</sup>, Rebecca L. McCulley<sup>29</sup>, Brett A. Melbourne<sup>19</sup>, Charles E. Mitchell<sup>22</sup>, Joslin L. Moore<sup>30</sup>, Brent Mortensen<sup>3</sup>, Lydia R. O'Halloran<sup>31</sup>, John L. Orrock<sup>18</sup>, Jesús Pascual<sup>6</sup>, Suzanne M. Prober<sup>32</sup>, David A. Pyke<sup>20</sup>, Anita C. Risch<sup>33</sup>, Martin Schuetz<sup>33</sup>, Melinda D. Smith<sup>10</sup>, Carly J. Stevens<sup>34</sup>, Lauren L. Sullivan<sup>3</sup>, Ryan J. Williams<sup>3</sup>, Peter D. Wragg<sup>1</sup>, Justin P. Wright<sup>35</sup> & Louie H. Yang<sup>36</sup>

**Human alterations to nutrient cycles<sup>1,2</sup> and herbivore communities<sup>3-7</sup> are affecting global biodiversity dramatically<sup>2</sup>. Ecological theory predicts these changes should be strongly counteractive: nutrient addition drives plant species loss through intensified competition for light, whereas herbivores prevent competitive exclusion by increasing ground-level light, particularly in productive systems<sup>8,9</sup>. Here we use experimental data spanning a globally relevant range of conditions to test the hypothesis that herbaceous plant species losses caused by eutrophication may be offset by increased light availability due to herbivory. This experiment, replicated in 40 grasslands on 6 continents, demonstrates that nutrients and herbivores can serve as counteracting forces to control local plant diversity through light limitation, independent of site productivity, soil nitrogen, herbivore type and climate. Nutrient addition consistently reduced local diversity through light limitation, and herbivory rescued diversity at sites where it alleviated light limitation. Thus, species loss from anthropogenic eutrophication can be ameliorated in grasslands where herbivory increases ground-level light.**

The diversity of life on Earth underlies critical ecosystem functions and economically important services<sup>10</sup>, and the current rapid rate of biodiversity loss<sup>2</sup> lends urgency to the task of understanding the forces that maintain biodiversity. Resources required for economic growth, energy and agriculture have all impacted natural ecosystems on a global scale. Introductions and extirpations of herbivore species, particularly as land is converted for grazing<sup>3-7</sup>, and increased nutrient supply are symptoms of humanity's global footprint<sup>1,2</sup>. Such widespread alteration of herbivores and nutrient supply may jointly determine the future diversity of ecosystems. For example, in highly productive, eutrophic

systems in which plant species extinction is likely to occur owing to a loss of ecological niches<sup>8</sup>, ecological theory predicts that herbivores can act to maintain local-scale plant diversity if they selectively consume the superior resource competitors<sup>9</sup>. Empirical studies in many ecosystem types find highly variable effects of herbivores on plant species diversity<sup>11-13</sup>, with the magnitude of herbivore mediation of diversity frequently observed to be greatest in regions of high ecosystem productivity<sup>11-17</sup>. Thus, ecosystem productivity and its regional climate drivers have been observed to mediate the local-scale effects of herbivores on plant diversity. The availability of ground-level light is the commonly postulated mechanism modulating the relationships among plant diversity, herbivory and observed gradients of plant productivity. However, these patterns and predictions have primarily emerged from studies across observed gradients of productivity or reviews and meta-analyses based on an extremely limited number of single-site experiments that manipulate both nutrients and herbivory, often with different methods<sup>11-15,17</sup>. In most of these studies, ground-level light has not been measured. Thus, the generality of these effects is only suggestive, and the mechanisms underlying the observed relationships remain elusive.

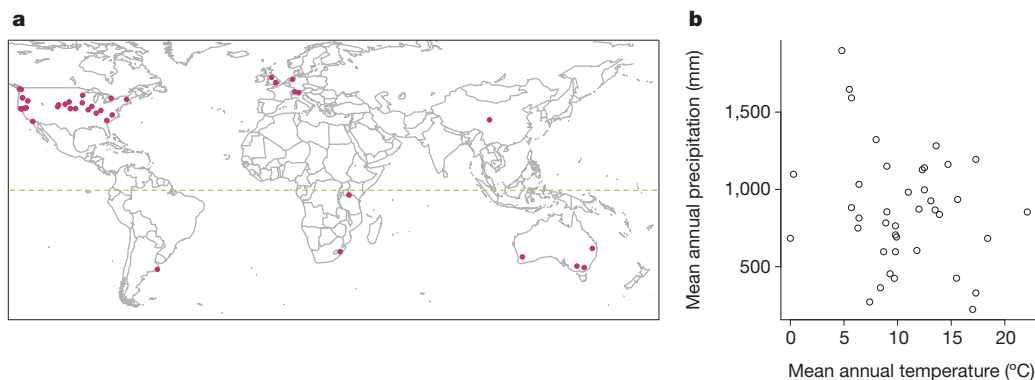
Local-scale plant diversity is likely to be maintained through an interdependent system of interactions with multiple plant species that are consumed by the same herbivores and that compete for light and nutrients. In particular, terrestrial plants compete for nutrients and light at the scale of interactions among individuals (neighbourhood of  $\sim 1 \text{ m}^2$  in grasslands), and one important mechanism for maintaining local coexistence is a trade-off in competitive ability for nutrients (below-ground) and for carbon via light (above-ground)<sup>18-22</sup>. Nutrient enrichment can lead to competitive exclusion of inferior competitors for light<sup>19,20</sup>,

<sup>1</sup>Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA. <sup>2</sup>Department of Entomology, University of Maryland, College Park, Maryland 20742, USA.

<sup>3</sup>Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA. <sup>4</sup>Institute for Chemistry and Biology of the Marine Environment, Carl-von-Ossietzky University, 26382 Wilhelmshaven, Oldenburg, Germany. <sup>5</sup>Department of Wildland Resources and the Ecology Center, Utah State University, Logan, Utah 84322, USA. <sup>6</sup>Instituto de Investigaciones Marinas y Costeras (IIIMyC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Mar del Plata 7600, Argentina. <sup>7</sup>Department of Biology, Wake Forest University, Winston-Salem, North Carolina 27109, USA.

<sup>8</sup>School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>9</sup>Agricultural Research Service (ARS), United States Department of Agriculture, Fort Collins, Colorado 80526, USA. <sup>10</sup>Department of Forest, Rangeland and Watershed Stewardship, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>11</sup>Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824, USA. <sup>12</sup>ARC Centre of Excellence for Environmental Decisions, School of Biological Sciences, The University of Queensland, Queensland 4072, Australia. <sup>13</sup>School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland. <sup>14</sup>Department of Ecology and Evolutionary Biology, University of Toronto Scarborough, Ontario M1C 1A4, Canada. <sup>15</sup>State Key Laboratory of Grassland and Agro-Ecosystems, Research Station of Alpine Meadow and Wetland Ecosystems, School of Life Sciences, Lanzhou University, Lanzhou, 730000 Gansu, China. <sup>16</sup>Division of Biological Sciences, University of California, San Diego, California 92093, USA. <sup>17</sup>Department of Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, UK. <sup>18</sup>Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>19</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder Colorado 80309, USA. <sup>20</sup>US Geological Survey, Forest and Rangeland Ecosystem Science Center, Corvallis, Oregon 97331, USA. <sup>21</sup>Queensland University of Technology, Biogeosciences, Brisbane, Queensland 4001, Australia. <sup>22</sup>Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. <sup>23</sup>Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK. <sup>24</sup>School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>25</sup>School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588, USA. <sup>26</sup>Berkeley Initiative for Global Change Biology, University of California, Berkeley 94704, USA. <sup>27</sup>Department of Plant Biology, University of Illinois at Urbana-Champaign, Illinois 61820, USA. <sup>28</sup>Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. <sup>29</sup>Department of Plant & Soil Sciences, University of Kentucky, Lexington, Kentucky 40546, USA. <sup>30</sup>Australian Research Center for Urban Ecology, c/o School of Botany, University of Melbourne, Victoria 3010, Australia, and School of Biological Sciences, Monash University, Victoria 3800, Australia. <sup>31</sup>Department of Zoology, Oregon State University, Corvallis, Oregon 97331, USA.

<sup>32</sup>CSIRO Ecosystem Sciences, Wembley, West Australia 6913, Australia. <sup>33</sup>Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf 8903, Switzerland. <sup>34</sup>Lancaster Environment Center, Lancaster University, Lancaster LA1 4YQ, UK. <sup>35</sup>Department of Biology, Duke University, Durham, North Carolina 27708, USA. <sup>36</sup>Department of Entomology, University of California, Davis, California 95616, USA.



**Figure 1 | Geographic and climatic distribution of experimental sites.**

**a**, Locations of the 40 'Nutrient Network' sites at which the full factorial experiment was replicated. At each site, nutrient supply and plot access by large herbivores were manipulated. Sites contributed 1–4 years of plot-scale

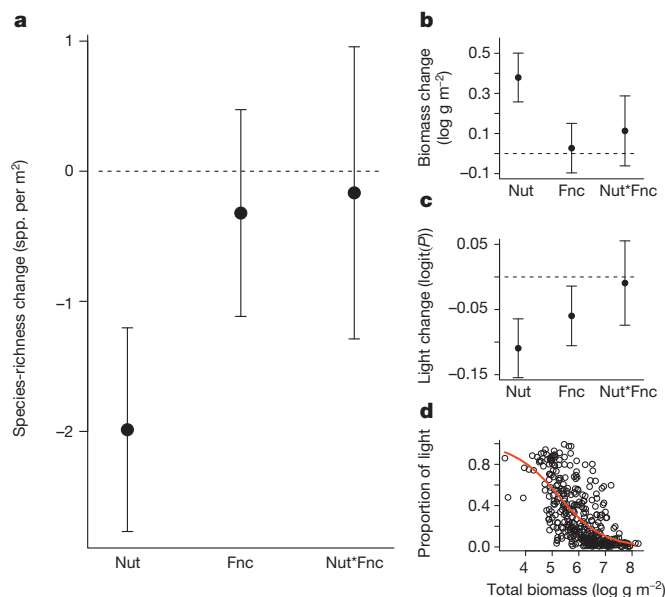
but herbivores can remove plant biomass, potentially alleviating understorey light limitation. However, herbivory creates another axis of potential trade-offs among plant species, involving investment in rapid growth and light capture versus investment in defence against herbivory<sup>23–26</sup>. These interactions result in a dynamic local community in which composition responds quickly to changes in the strength of nutrient limitation or herbivory<sup>25,26</sup>. In eutrophied systems, where nutrient limitation is alleviated and productivity is increased, theory predicts that these tradeoffs among plant strategies will simplify the plant community to species sharing an herbivore and competing for a single resource—light<sup>9,19</sup>.

We used an experiment replicated at 40 sites on 6 continents (Fig. 1) to test the hypothesis that herbivores mediate species losses caused by nutrient addition by increasing ground-level light, particularly in eutrophic and highly productive systems. We manipulated herbivores and nutrients using a factorial experiment (nutrient addition  $\times$  exclusion of herbivores  $>$  about 50 g, see Methods and ref. 27 for details) replicated in 40 sites dominated by herbaceous plants, spanning broad environmental gradients of productivity (114 to 1,976 g m<sup>-2</sup> yr<sup>-1</sup>), precipitation (mean annual precipitation from 224 to 1,898 mm yr<sup>-1</sup>), temperature (mean annual temperature from 0 to 22.1 °C), and soil nitrogen (mean percentage of soil N from 0.018 to 1.182%) (Fig. 1 and Extended Data Table 1). In each plot, we measured local-scale responses of productivity, light and the number of plant species (diversity) using standard methods<sup>27</sup>. We also examined site-level covariates including precipitation, temperature, herbivory intensity, soil nitrogen and atmospheric-nitrogen deposition rates. Although most sites provided 3 years of data, a subset of sites contributed 4 years of post-treatment data, and a few sites, established later, provided only 1 or 2 years of data (Extended Data Table 1). Effects of the experimental treatments were broadly consistent across all years of treatments (Extended Data Fig. 1); we present results from the three year duration in the main text, for a balance of spatial and temporal extent (see Extended Data Tables 2–8 for statistical models that describe responses in data collected 3 years after initiation of the experimental treatments).

Our results support an important mechanism by which nutrients lead to diversity loss. In particular, nutrient addition caused declines in diversity (Fig. 2a and Extended Data Table 2,  $P < 0.001$ ), increased total plant biomass (Fig. 2b and Extended Data Table 3,  $P < 0.001$ ), and increased light limitation (reduced transmission of photosynthetically active radiation (PAR) to ground level, Fig. 2c and Extended Data Table 4,  $P < 0.001$ ) both inside and outside of fences. Ground-level light availability, a function of light interception by live and dead biomass, declined with increasing total biomass (Fig. 2d and Extended Data Table 5,  $P < 0.001$ ). This result is consistent with eutrophication-induced loss of niches for coexistence<sup>8,9,19,28</sup> and demonstrates the generality of eutrophication as a primary force controlling the diversity of grassland communities by reducing ground-level light<sup>20</sup>.

photosynthetically active radiation (PAR), plant species richness and total plant biomass data. **b**, Study sites represent a wide range of mean annual temperature and precipitation ( $n = 40$ ). Additional site details are provided in Extended Data Table 1.

Although the removal of vertebrate herbivores did not have consistent effects on diversity (Fig. 2a,  $P = 0.522$ ) or biomass (Fig. 2b,  $P = 0.803$ ), herbivore removal increased light limitation (Fig. 2c,  $P = 0.013$ ). The lack of a consistent effect of herbivore removal on diversity across these globally distributed grassland sites (Fig. 2a) reflects the broad range of positive and negative effects found in past studies<sup>11,17,29</sup>. However, a critical assumption underlying the hypothesis that grassland diversity is jointly controlled by nutrient supply and consumers is that diversity should be rescued consistently by herbivory. In both ambient and eutrophied plots, herbivory should lead to greater diversity because herbivores can alleviate ground-level light limitation, thereby increasing the number of possible tradeoffs (nutrients, light) that maintain plant species diversity<sup>9,16</sup>.



**Figure 2 | Mixed-effects model parameters showing average response of plots ( $n = 360$ ) to 3 years of nutrient addition and herbivore exclusion by fencing. a–d**, Nutrient addition (Nut) and herbivore exclusion by fencing (Fnc) represent the difference from control plots; Nut\*Fnc is the additional effect of combining nutrients and fences (the interaction). Error bars represent the 95% confidence interval. **a**, Plot-scale diversity declines with nutrients, but is not consistently altered with fencing. **b**, Total biomass increases with nutrients, but is not consistently affected by fencing. **c**, Exclusion of herbivores and addition of nutrients independently reduce ground-level light. **d**, The proportion of light reaching the ground declines with increasing aboveground biomass.

We tested whether the inconsistent herbivore effects on plant diversity reflected variable herbivore effects on light and found that plant diversity increased quantitatively with herbivore effects on ground-level light (Fig. 3,  $P = 0.003$ ); nutrient addition did not modify this relationship (Extended Data Fig. 2). Sites with the greatest effects of herbivores on light and diversity spanned four continents and were dominated by larger vertebrates including wild and domestic ungulates, macropods, and lagomorphs (Fig. 3 and Extended Data Table 1). Thus, we show that under conditions of experimentally imposed nutrient supply gradients and among sites spanning greater than a 26-fold observed productivity gradient, herbivores rescue plant diversity to the extent that they enhance ground-level light, regardless of herbivore identity or nutrient supply.

Herbivore effects on plant diversity were not related to variation in soil nitrogen, nitrogen deposition rates, or site productivity. The change in ground-level light caused by removing herbivores was greatest at sites with high herbivory intensity (estimated as change in biomass in response to fencing;  $P = 0.006$ , AIC (Akaike information criterion)-weighted importance = 0.98, Extended Data Table 6). Herbivory intensity, in turn, was greatest at sites with a cool dry-season climate ( $P = 0.01$ , importance = 1.0, Extended Data Table 7) and sites at which the annual temperature is relatively warm ( $P = 0.03$ , importance = 0.52) and constant ( $P = 0.05$ , importance = 0.63). However, the change in diversity due to herbivores was best described by their effect on ground-level light ( $P = 0.012$ , importance = 1.0, Extended Data Table 8); site-level climate, productivity, soil nitrogen, nitrogen deposition rates, and herbivory intensity were not significant descriptors of changes in site-level plant diversity ( $P > 0.05$  and importance  $< 0.25$  for these factors). Thus, climate, which predicts herbivory intensity, places an ultimate constraint on the effects of herbivores on plant diversity, but local plant diversity is determined primarily through herbivore effects on ground-level light. These experimental data demonstrate that across a wide

range of the world's grasslands, herbivores serve as an important force to maintain plant diversity in grassland ecosystems in which they increase ground-level light availability, consistent with the prediction that light limitation is a critical factor controlling grassland species diversity<sup>9,16</sup>, but counter to the interpretation of nutrient supply or ecosystem productivity as the dominant force constraining herbivore effects on local plant diversity<sup>11–17</sup>.

A predictive understanding of the forces controlling grassland diversity is critical for informing issues of environmental and agricultural sustainability on all continents because of the steady conversion of the world's grasslands for use in livestock production<sup>3–7</sup>. Previous work observed that herbivores have the greatest effects on diversity in high productivity ecosystems<sup>11–16</sup>, whereas the experimental results presented here demonstrate that in grasslands in which herbivores increase ground-level light, they rescue plant diversity regardless of nutrient addition or environmental productivity. This result is consistent with ecological theory<sup>9,16,20</sup>, simultaneously providing greater mechanistic understanding<sup>20</sup> and clarifying the apparent overall lack of response of plant diversity to herbivory<sup>11</sup>. The results from our global-scale experiment suggest that in cases in which anthropogenic nutrient inputs to natural systems are high, grassland plant diversity will decline. However, in grasslands in which herbivory leads to increased ground-level light availability, we expect that these diversity losses will be ameliorated.

## METHODS SUMMARY

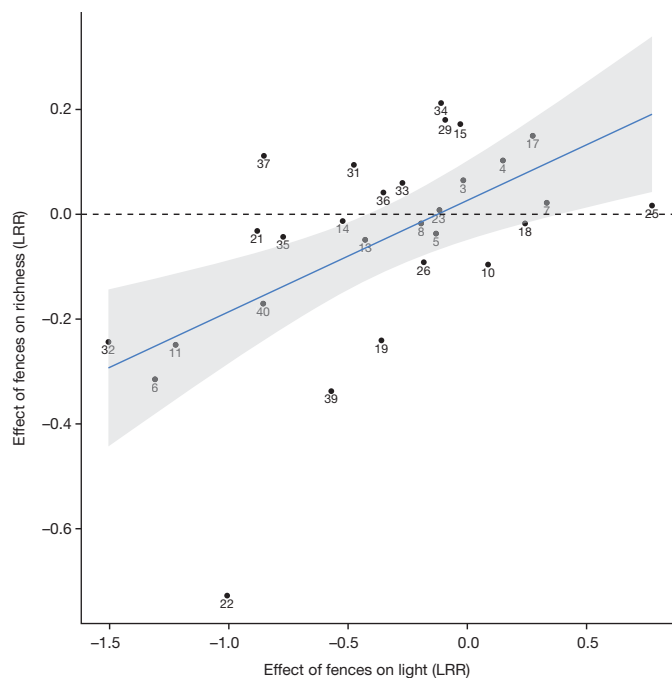
All 40 herbaceous-dominated ('grassland') sites in the analysis (Fig. 1) implemented a full factorial combination of nutrient addition ('control' or 'all nutrients') and herbivore exclusion ('control' or 'fenced'). The experimental design, treatments and sampling procedures to document plant diversity, biomass, light interception by the canopy, and soil chemistry were replicated at all sites, as detailed in ref. 27 and described in the Methods. Climate data were derived for all sites using the WorldClim database (version 1.4)<sup>30</sup>. All sites contributed at least 1 year of post-treatment data. Light, biomass and species richness were measured concurrently at 29 sites contributing 3 or more years of data (Extended Data Table 1); we focus on these in our main analyses.

We developed mixed-effects models with 'site' and 'block nested within site' as random effects using R (version 3.1; R Foundation for Statistical Computing). Analyses in Fig. 2a, b were performed using the nlme R library; we used the lme4 R library when the proportion of light (binomial error structure) was the response variable (Fig. 2c, d). We used the glm R function to analyse changes in each factor estimated as  $\log(S_{f+}/S_{f-})$ , where  $S_{f+}$  is species richness or proportion PAR reaching the ground in fenced plots and  $S_{f-}$  represents the comparable control plot measurement (Fig. 3). Finally, we used the dredge function in the MuMIn R library to assess the relative importance of potentially covarying site-level factors. We fit all possible models, deriving parameter values with error, and kept those models not more than four AIC<sub>C</sub> (AIC corrected for small sample size) units higher than the best model. For each parameter, we further estimated AIC-weighted 'importance', a unitless metric indicating summed Akaike information across all models in which the parameter appears. Importance ranges from 0 (parameter not given explanatory weight) to 1 (parameter in all top models).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2013; accepted 11 February 2014.

Published online 9 March 2014.



**Figure 3 | Effects of herbivore exclusion by fencing on mean grassland species richness and the mean proportion of PAR reaching ground level at 29 sites after 3 years of treatment.** Log response ratios (LRRs) compare light and richness inside and outside fences. The grey region indicates the 95% confidence interval around the regression ( $P = 0.003$ ). Extended Data Table 1 shows site number codes. Herbivore exclusion generally leads to reduced ground-level light ( $< 0$  on x axis) coupled with reduced grassland species richness ( $< 0$  on y axis). Herbivore effects are consistent across fertilized and unfertilized plots (Extended Data Fig. 2).

1. Foley, J. A., Monfreda, C., Ramankutty, N. & Zaks, D. Our share of the planetary pie. *Proc. Natl Acad. Sci. USA* **104**, 12585–12586 (2007).
2. Rockström, J. *et al.* A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
3. Gibson, D. *Grasses and Grassland Ecology* (Oxford Univ. Press, 2009).
4. Neely, C., Bunning, S. & Wilkes, A. *Review of Evidence on Drylands Pastoral Systems and Climate Change: Implications and Opportunities for Mitigation and Adaptation* (Food and Agriculture Organization of the United Nations, 2009).
5. Foley, J. A. *et al.* Solutions for a cultivated planet. *Nature* **478**, 337–342 (2011).
6. Millennium Ecosystem Assessment Panel. *Ecosystems and Human Well-being: Synthesis* (Island Press, 2005).
7. Wassenaar, T. *et al.* Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. *Glob. Environ. Change* **17**, 86–104 (2007).



8. Harpole, W. S. & Tilman, D. Grassland species loss resulting from reduced niche dimension. *Nature* **446**, 791–793 (2007).
9. Holt, R. D., Grover, J. & Tilman, D. Simple rules for interspecific dominance in systems with exploitative and apparent competition. *Am. Nat.* **144**, 741–771 (1994).
10. Díaz, S., Fargione, J., Chapin, F. S. & Tilman, D. Biodiversity loss threatens human well-being. *PLoS Biol.* **4**, e277 (2006).
11. Hillebrand, H. *et al.* Consumer versus resource control of producer diversity depends on ecosystem type and producer community structure. *Proc. Natl Acad. Sci. USA* **104**, 10904–10909 (2007).
12. Proulx, M. & Mazumder, A. Reversal of grazing impact on plant species richness in nutrient-poor vs. nutrient-rich ecosystems. *Ecology* **79**, 2581–2592 (1998).
13. Worm, B., Lotze, H. K., Hillebrand, H. & Sommer, U. Consumer versus resource control of species diversity and ecosystem functioning. *Nature* **417**, 848–851 (2002).
14. Olff, H. & Ritchie, M. E. Effects of herbivores on grassland plant diversity. *Trends Ecol. Evol.* **13**, 261–265 (1998).
15. Milchunas, D. G. & Lauenroth, W. K. Quantitative effects of grazing on vegetation and soils over a global range of environments. *Ecol. Monogr.* **63**, 327–366 (1993).
16. Huisman, J. & Weissing, F. J. Light-limited growth and competition for light in well-mixed aquatic environments: an elementary model. *Ecology* **75**, 507–520 (1994).
17. Bakker, E. S., Ritchie, M. E., Olff, H., Milchunas, D. G. & Knops, J. M. H. Herbivore impact on grassland plant diversity depends on habitat productivity and herbivore size. *Ecol. Lett.* **9**, 780–788 (2006).
18. Weissing, F. J. & Huisman, J. Growth and competition in a light gradient. *J. Theor. Biol.* **168**, 323–336 (1994).
19. Dybzinski, R. & Tilman, D. Resource use patterns predict long-term outcomes of plant competition for nutrients and light. *Am. Nat.* **170**, 305–318 (2007).
20. Hautier, Y., Niklaus, P. A. & Hector, A. Competition for light causes plant biodiversity loss after eutrophication. *Science* **324**, 636–638 (2009).
21. Newman, E. I. Competition and diversity in herbaceous vegetation. *Nature* **244**, 310 (1973).
22. Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
23. Coley, P. D., Bryant, J. P. & Chapin, F. S. Resource availability and plant antiherbivore defense. *Science* **230**, 895–899 (1985).
24. Strauss, S. Y., Rudgers, J. A., Lau, J. A. & Irwin, R. E. Direct and ecological costs of resistance to herbivory. *Trends Ecol. Evol.* **17**, 278–285 (2002).
25. Lind, E. M. *et al.* Life-history constraints in grassland plant species: a growth-defence trade-off is the norm. *Ecol. Lett.* **16**, 513–521 (2013).
26. Grime, J. P. & Pierce, S. *The evolutionary strategies that shape ecosystems* (Wiley-Blackwell, 2012).
27. Borer, E. T. *et al.* Finding generality in ecology: a model for globally distributed experiments. *Methods Ecol. Evol.* **5**, 65–73 (2014).
28. De Schrijver, A. *et al.* Cumulative nitrogen input drives species loss in terrestrial ecosystems. *Glob. Ecol. Biogeogr.* **20**, 803–816 (2011).
29. Crawley, M. J. *Plant ecology* (Blackwell Science, 1997).
30. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work uses data from the Nutrient Network (<http://nutnet.org>) experiment, funded at the site scale by individual researchers. Coordination and data management are supported by funding to E. Borer and E. Seabloom from the NSF Research Coordination Network (NSF-DEB-1042132) and Long Term Ecological Research (NSF-DEB-1234162 to Cedar Creek LTER) programs and the UMN Institute on the Environment (DG-0001-13). The Minnesota Supercomputer Institute hosts project data. We are grateful to F. Isbell for suggestions that improved the manuscript. Any use of trade names is for descriptive purposes only and does not imply endorsement by the US Government.

**Author Contributions** E.T.B., E.W.S., W.S.H. and E.M.L. are Nutrient Network coordinators. E.T.B., W.S.H., H.H. and D.S.G. developed and framed the research questions in this paper. All authors contributed data from this experiment. E.T.B. and E.W.S. analysed the data. D.S.G., W.S.H. and E.M.L. contributed to data analyses. E.T.B. wrote the paper with input from all authors. Supplementary Information Appendix S2 provides further information on author contributions.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.T.B. ([borer@umn.edu](mailto:borer@umn.edu)).

## METHODS

**Site selection.** The Nutrient Network (NutNet) is a network of researchers working at herbaceous-dominated ('grassland') sites in countries spread across six continents, who are performing coordinated, globally distributed observations and experiments. The full experimental design is detailed here and in ref. 27. All NutNet sites are located in areas dominated by herbaceous vegetation representing the regional species composition (for example, shortgrass steppe, tallgrass prairie), referred to as 'grassland' here. The NutNet experimental design analysed here is a completely randomized block design with four  $5 \times 5$  m plots per block and three replicate blocks at most sites (with blocks ranging from 1 ( $n = 1$  site) to 5 ( $n = 3$  sites)). Within-site replication is used to determine relative strength of responses, but the main experimental replication comes from the number of sites.

**Experimental treatments.** All 40 sites included in the current analysis (Fig. 1) implemented a full factorial combination of nutrient addition ('control' or 'all nutrients') and consumer density ('control' or 'fenced') for a total of four treatments in randomized, complete blocks. Standard nutrient addition and sampling protocols were carefully replicated among sites<sup>27</sup>. All sites collected data before application of treatments (year 0); most sites began sampling in 2007, but a subset began sampling in subsequent years. Nutrient and fencing treatments<sup>27</sup> were implemented the following year (year 1) and have been maintained continuously since then. All sites contributed at least 1 year of post-treatment data; 39 of these sites contributed 3 or more years of post-treatment data. Light, biomass and species richness measurements (see below) were conducted concurrently at 29 sites contributing 3 or more years of data; we focus on these in our main analyses.

Fences designed to exclude aboveground mammalian herbivores (> about 50 g) were erected around two plots in each block, one receiving a nutrient combination (described next) and one used as an ambient nutrient control plot. Fences were 230 cm tall with the lower 90 cm surrounded by 1-cm woven wire mesh. An additional 30-cm outward-facing flange was stapled to the ground to exclude digging animals (for example, rabbits, voles), although not fully subterranean animals (for example, gophers, moles). Four strands of barbless wire were strung at equal vertical distances above the wire mesh. Enclosures were built at all sites before the second year of plant growth. Although most (33) sites built fences exactly to these specifications, a few sites (8) faced challenges (for example, snow-pack, materials availability, elephant activity) that required minor modifications. Modifications are described in Supplementary Table 1.

Nitrogen, phosphorus and potassium were applied annually to experimental plots; micronutrients were applied once at the start of the experiment to avoid toxic levels from over-application. Nutrient addition rates and sources were: 10 g N m<sup>-2</sup> yr<sup>-1</sup> as timed-release urea ((NH<sub>2</sub>)<sub>2</sub>CO), 10 g P m<sup>-2</sup> yr<sup>-1</sup> as triple-super phosphate (Ca(H<sub>2</sub>PO<sub>4</sub>)<sub>2</sub>), 10 g K m<sup>-2</sup> yr<sup>-1</sup> as potassium sulphate (K<sub>2</sub>SO<sub>4</sub>) and 100 g m<sup>-2</sup> yr<sup>-1</sup> of a micronutrient mix of Fe (15%), S (14%), Mg (1.5%), Mn (2.5%), Cu (1%), Zn (1%), B (0.2%) and Mo (0.05%).

Each sampling area was separated by at least 1.5 m from neighbouring plots (1 m walkway and 0.5 m within-plot buffer), which served to minimize indirect effects of treatments in one plot on adjacent plots (for example, nutrient leaching, shading or mycelial networks). Note that the nutrient and fence treatments had strong measurable effects on plant responses (for example, biomass, richness) indicating that plots and measurements were sufficiently sized and spaced.

**Species diversity.** All NutNet sites followed standard sampling protocols. A randomly designated 1 × 1 m subplot within each 5 × 5 m plot was permanently marked and sampled annually at peak biomass. In the 1 × 1 m permanently marked subplot, cover was estimated visually to the nearest 1% for every species overhanging the subplot; cover estimates also included woody over-story, litter, bare soil and rock.

**Productivity.** Adjacent to the permanent 1 × 1 m cover subplot, standing crop was estimated destructively by clipping at ground level all aboveground biomass of individual plants rooted within two 1 × 0.1 m strips (for a total of 0.2 m<sup>2</sup>). All biomass was dried at 60 °C to constant mass before weighing to the nearest 0.01 g. Weights were multiplied by five to estimate grams per square meter. Pre-treatment data (year 0) from each site in this study demonstrate high correlation 0.976 (95% confidence interval: 0.955–0.987) between year 0 mean plant biomass in the control plots ( $n = 3$  for most sites) and year 0 mean plant biomass for the site as characterized by all plots ( $n = 30$  for most sites).

**Light interception.** At the time of biomass clipping, photosynthetically active radiation (PAR,  $\mu\text{mol photons per m}^2 \text{ per s}$ ) was determined at approximately solar noon (between 11:00 and 14:00). Two measurements, integrated across a 1-m light ceptometer, were made at ground level from opposite corners of each 1-m<sup>2</sup> plant-diversity plot, diagonal to each other, and one measurement was made above

the canopy of each plot. We calculated the proportion of PAR available at ground level as the ratio of the average of the ground level to the ambient measurements. **Climate.** We used the WorldClim database to derive comparable climate data for all sites (version 1.4; <http://www.worldclim.org/bioclim>). This database provides high-resolution interpolated global climate data for stations with 10 to 30 years of data<sup>30</sup>. To examine climate covariates with site-level fencing effects on net consumption (biomass inside minus outside of fences), light and richness, we used climate variables that summarized the mean and seasonality of site-level temperature and precipitation. These were (BIO designator indicates the variable code in the WorldClim database): mean annual temperature (°C; BIO1), mean maximum temperature of the warmest month (°C; BIO5), mean minimum temperature of the warmest month (°C; BIO5), mean annual precipitation (mm per year; BIO12), precipitation variability (coefficient of variation in precipitation among months; BIO15), rainfall-potential evapotranspiration (mm per month), temperature variability (standard deviation of temperature among months; BIO4), mean temperature in the wettest quarter (°C; BIO8), and mean temperature in the driest quarter (°C; BIO9).

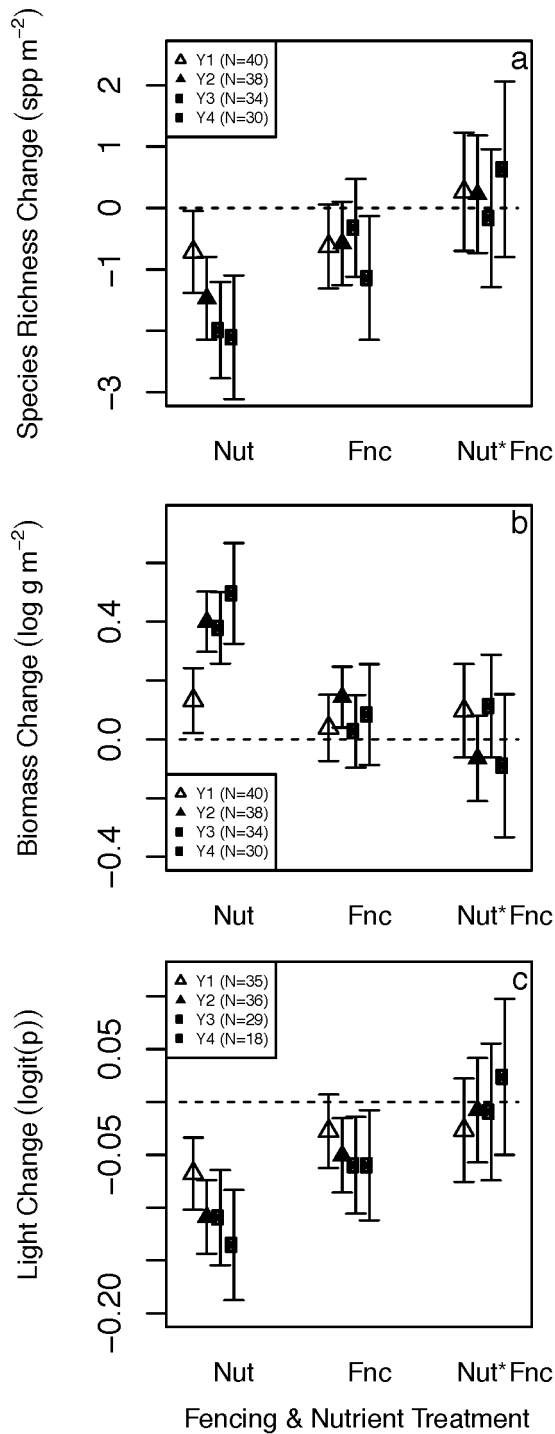
**Nitrogen deposition.** We used modelled nitrogen deposition rates<sup>31</sup> to determine the annual atmospheric N deposition (kg N ha<sup>-1</sup> yr<sup>-1</sup>) for each experimental site (associated with model output based on latitude and longitude). N deposition was modelled based on existing measurements and future projections using a global three-dimensional chemistry-transport model (TM3)<sup>31</sup>. The spatial resolution of the model, 5 degrees longitude by 3.75 degrees latitude, and the resolution of the output grid (50 × 50 km sub-grids), provide sufficient resolution to distinguish site-level variation in annual N deposition among our experimental sites.

**Statistical analysis.** To explore the independent and interactive effects of vertebrate herbivory and nutrient supply on species diversity, total biomass and photosynthetically active radiation, we developed mixed effects models with site and block within site as random effects. Analyses in Fig. 2a, b were performed using the nlme library in R (R version 3.1; R Foundation for Statistical Computing); for Fig. 2c, d, we used the lme4 R library to fit models in which proportion of light was the response variable (binomial error structure and a proportion bounded between 0 and 1). Although not presented here, models using logit and arcsin square root transformations of the data generated qualitatively identical results. Site and block nested within sites were included in all regressions. We also estimated the effects of herbivores on richness and light at each site as the change in these factors resulting from fencing in both fertilized and unfertilized plots. Change in each factor was estimated as the log ratio of the treatment divided by the control,  $\log(S_{f+}/S_{f-})$ , where  $S_{f+}$  is the species richness or proportion PAR reaching the ground in fenced plots and  $S_{f-}$  is the species richness or proportion PAR reaching the ground in control plots. We examined residuals to ensure homogeneity of variance. Because of missing PAR data for a few sites, this analysis included 29 sites. The relationships were independent of whether plots had been fertilized (see Extended Data Fig. 2 for more details), so we present a final model of site means including both fertilized and unfertilized plots in the main text. The log ratio analyses were performed using the glm library in R (R version 3.1; R Foundation for Statistical Computing). We generated models separately for each experimental year (year 1,  $n = 40$ ; year 2,  $n = 38$ ; year 3,  $n = 34$ ; year 4,  $n = 30$ ). The results were broadly consistent (Extended Data Fig. 1), so we present results from 3 years of manipulations in the main text and results comparing 1 to 4 years of manipulations, greater spatial (year 1 and year 2) or temporal (year 4) extent, in Extended Data Fig. 1.

Finally, to examine the effects of climate and site productivity as predictors for site-level mean herbivore effects on biomass, ground-level light and plant richness, we analysed site-level mean values using model averaging<sup>32</sup>. The model averaging approach allowed us to assess the relative importance of a range of covarying factors, and to recognize explicitly that there could be a suite of similar models. Prior to fitting the models, all of the independent variables were standardized using the standardize function in the arm R library. 'Importance' in this modelling approach is a term representing the relativized sum of the Akaike weights summed across all of the models in which the parameter appears that are within four AIC<sub>C</sub> (AIC corrected for small sample size) units of the model with the lowest AIC. Importance ranges from 0 (parameter not given explanatory weight) to 1 (parameter in all top models). We used the dredge function in the MuMIn R library to fit all possible models. We estimated parameter values, errors, and AIC-weighted importance using the model.avg function in the MuMIn R library.

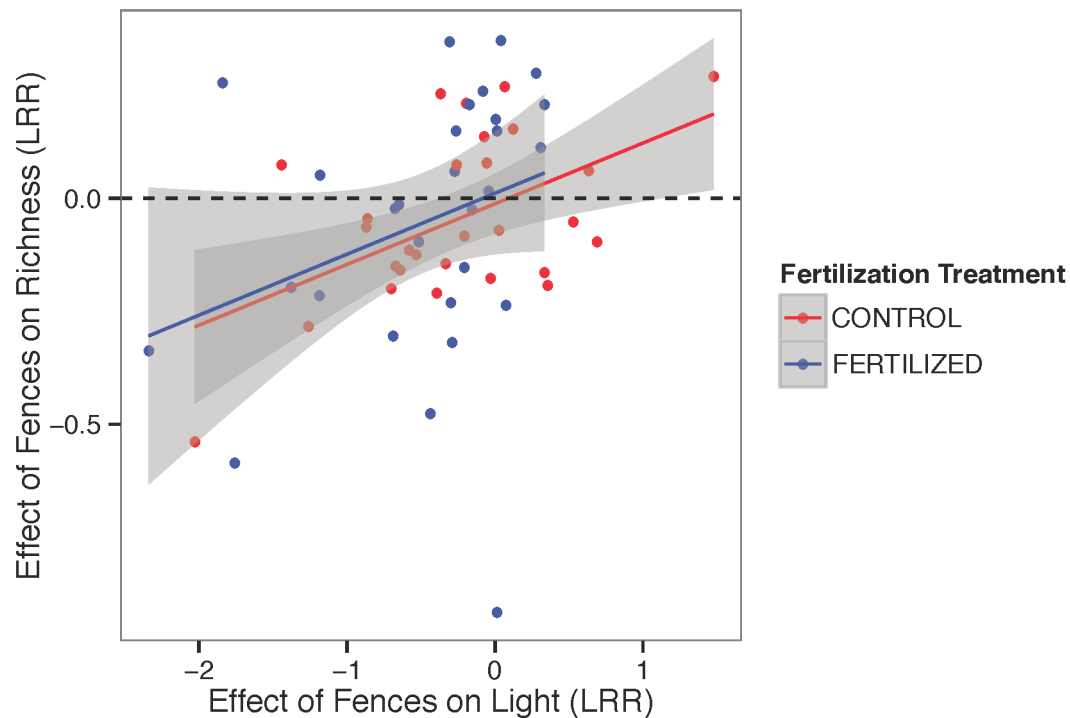
31. The Oak Ridge National Laboratory Distributed Active Archive Center. <http://daac.ornl.gov/> (accessed 25 August 2011).

32. Grueber, C. E., Nakagawa, S., Laws, R. J. & Jamieson, I. G. Multimodel inference in ecology and evolution: challenges and solutions. *J. Evol. Biol.* **24**, 699–711 (2011).



**Extended Data Figure 1 | Regression parameters for treatment effects.** a–c, All available data are shown for richness (a), total biomass (b) and ambient light reaching ground level (c). Error bars represent  $\pm 2$  s.e.m. Treatment years and their associated sample sizes are shown in each panel. One- and two-year models represent greater spatial extent and replication, but reduced temporal extent compared to Fig. 2 in the main text. Four-year models represent longer temporal effects, but reduced spatial extent, particularly for light measurements. All models were fitted as in Extended Data Tables 2–4 and described in the Methods.





**Extended Data Figure 2 | Fertilization does not alter the relationship between ‘fence’ effects on light and diversity.** The log response ratio (LRR) model of the effect of fences (herbivore exclusion) on richness and light (year 3 data) demonstrates no additional effect of nutrient addition on the relationship shown in Fig. 3. The grey region indicates the 95% confidence interval around

the regression. The effect of fences on ground-level light predicts changes in plot-scale species richness ( $P = 0.00254$ ), whereas fertilization is not included in the final statistical model of this relationship ( $P > 0.05$ ). Thus, the magnitude of the effect of grazers on richness is dependent on the magnitude of their effect on light regardless of whether a plot has been fertilized.

Extended Data Table 1 | Sites contributing experimental data

Site #	Site code*	Country	Continent	Lat.	Long.	Exp't years	MAT (°C)	MAP (mm)	Mean soil %N	Control plot mass (g m <sup>-2</sup> )	Control plot richness (m <sup>-2</sup> )
1	azi.cn	China	Asia	33.670	101.870	4	0.0	684	0.556	411.40	33.17
2	barta.us	USA	N Amer.	42.245	-99.652	4	8.7	597	0.057	442.90	13.75
3	bldr.us*	USA	N Amer.	39.972	-105.234	4	9.7	425	0.089	283.45	8.10
4	bnch.us*	USA	N Amer.	44.277	-121.968	4	5.5	1647	0.537	281.87	9.43
5	bogong.au*	Australia	Australia	-36.874	147.254	4	5.7	1592	0.722	554.27	18.93
6	burrawan.au*	Australia	Australia	-27.735	151.140	4	18.4	683	0.087	251.18	10.33
7	cbgb.us*	USA	N Amer.	41.785	-93.385	3	9.0	855	0.063	870.38	8.08
8	cdcr.us*	USA	N Amer.	45.401	-93.201	4	6.3	750	0.039	270.40	10.55
9	cdpt.us	USA	N Amer.	41.200	-101.630	4	9.3	454	0.097	233.59	11.72
10	cowi.ca*	Canada	N Amer.	48.460	-123.380	4	9.8	764	0.374	1018.63	4.80
11	elliott.us*	USA	N Amer.	32.875	-117.052	4	17.0	224	0.165	463.98	11.25
12	frue.ch	Switz.	Europe	47.113	8.542	3	9.0	1150	0.364	1046.54	13.58
13	gilb.za*	S Africa	Africa	-29.284	30.292	3	13.1	926	1.182	223.23	27.08
14	hall.us*	USA	N Amer.	36.872	-86.702	4	13.6	1282	0.139	948.55	6.13
15	hart.us*	USA	N Amer.	42.724	-119.498	4	7.4	272	0.109	113.82	10.33
16	hero.uk	UK	Europe	51.411	-0.639	4	9.9	692	NA	740.88	16.00
17	hnvr.us*	USA	N Amer.	43.419	-72.138	3	6.4	1033	0.395	524.58	9.50
18	hopl.us*	USA	N Amer.	39.013	-123.060	4	12.3	1127	NA	347.28	18.43
19	kiny.au*	AUS	Australia	-36.200	143.750	4	15.5	426	0.120	490.11	11.20
20	koffler.ca	Canada	N Amer.	44.024	-79.536	1	6.4	815	NA	964.47	9.39
21	konz.us*	USA	N Amer.	39.071	-96.583	4	12.0	872	NA	433.15	15.07
22	lancaster.uk*	UK	Europe	53.986	-2.628	3	8.0	1322	1.180	126.97	9.50
23	look.us*	USA	N Amer.	44.205	-122.128	4	4.8	1898	1.170	332.69	7.20
24	marc.ar	Argent.	S Amer.	-37.715	-57.425	1	13.9	838	NA	953.79	11.90
25	mcla.us*	USA	N Amer.	38.864	-122.406	4	13.5	867	NA	554.20	5.33
26	mtca.au*	AUS	Australia	-31.782	117.611	4	17.3	330	0.092	115.40	14.66
27	pape.de	Germany	Europe	53.086	7.473	4	8.9	783	0.190	1292.78	5.00
28	rook.uk	UK	Europe	51.406	-0.644	4	9.8	706	NA	233.19	10.92
29	sage.us*	USA	N Amer.	39.430	-120.240	4	5.7	882	0.714	159.92	12.36
30	saline.us	USA	N Amer.	39.050	-99.100	4	11.8	605	NA	490.44	10.75
31	sava.us*	USA	N Amer.	33.344	-81.651	4	17.3	1194	0.018	393.80	10.30
32	sereng.tz*	Tanzania	Africa	-2.255	34.513	3	22.1	854	0.133	362.90	12.58
33	sgs.us*	USA	N Amer.	40.817	-104.767	4	8.4	364	0.081	129.08	6.89
34	sier.us*	USA	N Amer.	39.236	-121.284	4	15.6	935	0.171	378.29	8.32
35	smith.us*	USA	N Amer.	48.207	-122.625	4	9.8	597	0.521	440.24	20.67
36	spin.us*	USA	N Amer.	38.136	-84.501	4	12.5	1140	0.261	425.16	9.53
37	trel.us*	USA	N Amer.	40.075	-88.829	3	11.0	982	0.273	1976.39	4.22
38	tyso.us	USA	N Amer.	38.519	-90.565	2	12.5	997	0.261	599.34	10.33
39	unc.us*	USA	N Amer.	36.008	-79.020	4	14.7	1162	0.162	427.24	11.73
40	valm.ch*	Switz.	Europe	46.631	10.372	4	0.3	1098	0.498	253.20	26.85

Nutrient Network sites with 1 to 4 years of experimental manipulations of both nutrient supply and herbivore access to plots. Control plot mass and control plot richness, the mean total biomass and mean number of species per m<sup>2</sup> in un-manipulated plots across all sample years; experiment years, the number of years of experimental data collected at each site for the current analyses; mean soil %N, the average percentage of nitrogen in soil of the site (across all plots, measured before establishment of experimental treatments); MAP is mean annual precipitation; and MAT is mean annual temperature. Lat. and Long. provide geographic coordinates of each site.

\*Sites used in the main text analyses (year 3 measurements of richness, biomass and photosynthetically active radiation (PAR)).

**Extended Data Table 2 | Statistical model for treatment effects on richness after 3 years of treatment ( $n = 29$ ) as a function of fertilization by N, P and K and micronutrients, herbivore exclusion by fencing, and their interaction**

	Value	Std.Error	DF	t-value	p-value
(Intercept)	11.965	1.085	264	11.031	<0.0001
fertFERTILIZED	-1.987	0.391	264	-5.075	<0.0001
excloseFENCE	-0.321	0.397	264	-0.809	0.4194
fertFERTILIZED:excloseFENCE	-0.166	0.561	264	-0.296	0.7677

Linear mixed-effects model was fit by maximum likelihood. Random effects in model were 'site' (s.d. = 5.60) and 'block within site' (s.d. = 1.22). The intercept is the estimated mean value of the control plots (no fence, no nutrients). This model is shown in Fig. 2a.



**Extended Data Table 3 | Statistical model for treatment effects on biomass after 3 years of treatment ( $n = 29$ ) as a function of fertilization by N, P and K and micronutrients, herbivore exclusion by fencing, and their interaction**

	Value	Std.Error	DF	t-value	p-value
(Intercept)	5.803	0.143	264	40.497	<0.0001
fertFERTILIZED	0.379	0.061	264	6.235	<0.0001
excloseFENCE	0.027	0.062	264	0.438	0.6619
fertFERTILIZED:excloseFENCE	0.113	0.087	264	1.296	0.1961

Linear mixed-effects model was fit by maximum likelihood. Random effects in model were 'site' (s.d. = 0.73) and 'block within site' (s.d. = 0.20). The intercept is the estimated mean value of the control plots (no fence, no nutrients). This model is shown in Fig. 2b.

**Extended Data Table 4 | Statistical model for treatment effects on proportion of photosynthetically active radiation (PAR) reaching ground level after 3 years of treatment ( $n = 29$ ) as a function of fertilization by N, P and K and micronutrients, herbivore exclusion by fencing, and their interaction**

	Value	Std.Error	DF	t-value	p-value
(Intercept)	0.399	0.046	264	8.615	<0.0001
fertFERTILIZED	-0.109	0.023	264	-4.850	<0.0001
excloseFENCE	-0.060	0.023	264	-2.614	0.0095
fertFERTILIZED:excloseFENCE	-0.009	0.032	264	-0.290	0.7721

Linear mixed-effects model was fit by maximum likelihood. Random effects in model were 'site' (s.d. = 0.23) and 'block within site' (s.d. =  $6.24 \times 10^{-6}$ ). The intercept is the estimated mean value of the control plots (no fence, no nutrients). This model is shown in Fig. 2c.

**Extended Data Table 5 | Statistical model for biomass effects on ground-level proportion of photosynthetically active radiation (PAR) after 3 years of treatment ( $n = 29$ ) as a function of total plot-scale biomass**

	Estimate	Std. Error	z-value	Pr(> z )
(Intercept)	-0.584	0.395	-1.479	0.1391
fertFERTILIZED	-1.108	0.361	-3.073	0.0021
excloseFENCE	-0.993	0.363	-2.736	0.0062
fertFERTILIZED:excloseFENCE	0.166	0.551	0.302	0.7626

Generalized linear mixed-effects model with logit link and binomial errors was fit by maximum likelihood. Random effects in model were 'site' (s.d. = 1.68) and 'block within site' (s.d. =  $1.06 \times 10^{-5}$ ). The intercept is the estimated mean value of the control plots (no fence, no nutrients). This model is shown in Fig. 2d.



**Extended Data Table 6 | Effects of climate, nitrogen deposition, soil nitrogen and site productivity on change in ground-level light across experimental fencing treatments after 3 years of treatment**

	Estimate	Std. Error	Adjusted SE	z value	p-value	Relative Importance
(Intercept)	-0.30123	0.07368	0.07827	3.849	<0.001	1.00
Change in mass across fence	-0.56329	0.19684	0.20635	2.730	0.006	0.98
Annual Precipitation Variability	0.29938	0.16493	0.17463	1.714	0.086	0.51
Wet Quarter Temperature	-0.42281	0.28903	0.29826	1.418	0.156	0.40
Mean Annual Temperature	-0.51396	0.46975	0.48080	1.069	0.285	0.36
Maximum Temperature	0.53099	0.35922	0.37327	1.423	0.155	0.35
Minimum Temperature	-0.28435	0.43556	0.44519	0.639	0.523	0.29
Dry Quarter Temperature	-0.14111	0.54170	0.55158	0.256	0.798	0.18
Annual Temperature Variability	0.12017	0.39288	0.40263	0.298	0.765	0.17
Annual N-Deposition Rate	-0.14619	0.18187	0.19180	0.762	0.446	0.08
Rainfall Potential Evapotranspir.	-0.06077	0.22652	0.23682	0.257	0.797	0.06
Net primary production in Y3	-0.05213	0.16746	0.17748	0.294	0.769	0.04

Summary results of change in site-level means of ground-level light after model averaging; all factors are normalized. Soil nitrogen was included in the original models, but was never significant so was dropped from final models because of missing values.

**Extended Data Table 7 | Effects of climate, nitrogen deposition, soil nitrogen and site productivity on site-level mean biomass change across experimental fencing treatments after 3 years of treatment**

	Estimate	Std. Error	Adjusted SE	z value	p-value	Relative Importance
(Intercept)	0.12111	0.04966	0.05247	2.308	0.021	1.00
Dry Quarter Temperature	-0.66911	0.26088	0.27013	2.477	0.013	1.00
Annual Temperature Variability	-0.58513	0.28798	0.29564	1.979	0.048	0.63
Annual Precipitation Variability	-0.10096	0.13851	0.14446	0.699	0.485	0.63
Minimum Temperature	0.65583	0.29851	0.30830	2.127	0.033	0.52
Wet Quarter Temperature	-0.28870	0.29658	0.30660	0.942	0.346	0.29
Mean Annual Temperature	0.34511	0.40053	0.40959	0.843	0.400	0.29
Net primary production in Y3	-0.11616	0.10543	0.11130	1.044	0.297	0.15
Mean Annual Precipitation	-0.01959	0.11602	0.12206	0.160	0.873	0.10
Rainfall Potential Evapotranspir.	-0.06785	0.16203	0.16898	0.402	0.688	0.09
Annual N-Deposition Rate	-0.09961	0.14467	0.15200	0.655	0.512	0.09

Summary results of site-level means of biomass off-take after model averaging; all factors are normalized. Soil nitrogen was included in the original models, but was never significant so was dropped from final models because of missing values.

**Extended Data Table 8 | Effects of climate, nitrogen deposition, soil nitrogen, site productivity, and change in light on change in site-level mean plant species richness across experimental fencing treatments after three years of treatments**

	Estimate	Std. Error	Adjusted SE	z value	p-value	Relative Importance
(Intercept)	-0.05642	0.03317	0.03511	1.607	0.108	1.00
Change in light across fence	0.21878	0.08325	0.08734	2.505	0.012	1.00
Minimum Temperature	-0.16634	0.20331	0.20831	0.798	0.425	0.21
Maximum Temperature	0.15296	0.14054	0.14489	1.056	0.291	0.19
Change in mass across fence	-0.0941	0.08964	0.09495	0.991	0.322	0.13
Annual Temperature Variability	-0.0571	0.26403	0.26864	0.213	0.832	0.13
Rainfall Potential Evapotranspir.	-0.2268	0.32134	0.32656	0.695	0.487	0.13
Mean Annual Temperature	-0.17278	0.18625	0.19128	0.903	0.366	0.11
Annual N-Deposition Rate	-0.063	0.08599	0.09065	0.695	0.487	0.08
Mean Annual Precipitation	0.18256	0.28651	0.29187	0.625	0.532	0.07
Annual Precipitation Variability	0.05612	0.07807	0.08242	0.681	0.496	0.06
Dry Quarter Temperature	0.05108	0.12828	0.13300	0.384	0.701	0.06
Wet Quarter Temperature	0.03975	0.07086	0.07498	0.530	0.596	0.04
Net primary production in Y3	0.04017	0.07414	0.07845	0.512	0.609	0.04

Summary results of change in site-level means of plant richness after model averaging; all factors are normalized. Soil nitrogen was included in the original models, but was never significant so was dropped from final models because of missing values.



# Eutrophication weakens stabilizing effects of diversity in natural grasslands

Yann Hautier<sup>1,2</sup>, Eric W. Seabloom<sup>1</sup>, Elizabeth T. Borer<sup>1</sup>, Peter B. Adler<sup>3</sup>, W. Stanley Harpole<sup>4</sup>, Helmut Hillebrand<sup>5</sup>, Eric M. Lind<sup>1</sup>, Andrew S. MacDougall<sup>6</sup>, Carly J. Stevens<sup>7</sup>, Jonathan D. Bakker<sup>8</sup>, Yvonne M. Buckley<sup>9,10</sup>, Chengjin Chu<sup>11</sup>, Scott L. Collins<sup>12</sup>, Pedro Daleo<sup>13</sup>, Ellen I. Damschen<sup>14</sup>, Kendi F. Davies<sup>15</sup>, Philip A. Fay<sup>16</sup>, Jennifer Firn<sup>17</sup>, Daniel S. Gruner<sup>18</sup>, Virginia L. Jin<sup>19</sup>, Julia A. Klein<sup>20</sup>, Johannes M. H. Knops<sup>21</sup>, Kimberly J. La Pierre<sup>22</sup>, Wei Li<sup>23</sup>, Rebecca L. McCulley<sup>24</sup>, Brett A. Melbourne<sup>15</sup>, Joslin L. Moore<sup>25,26</sup>, Lydia R. O'Halloran<sup>27</sup>, Suzanne M. Prober<sup>28</sup>, Anita C. Risch<sup>29</sup>, Mahesh Sankaran<sup>30,31</sup>, Martin Schuetz<sup>29</sup> & Andy Hector<sup>32</sup>

**Studies of experimental grassland communities<sup>1–7</sup> have demonstrated that plant diversity can stabilize productivity through species asynchrony, in which decreases in the biomass of some species are compensated for by increases in others<sup>1,2</sup>. However, it remains unknown whether these findings are relevant to natural ecosystems, especially those for which species diversity is threatened by anthropogenic global change<sup>8–11</sup>. Here we analyse diversity–stability relationships from 41 grasslands on five continents and examine how these relationships are affected by chronic fertilization, one of the strongest drivers of species loss globally<sup>8</sup>. Unmanipulated communities with more species had greater species asynchrony, resulting in more stable biomass production, generalizing a result from biodiversity experiments to real-world grasslands. However, fertilization weakened the positive effect of diversity on stability. Contrary to expectations, this was not due to species loss after eutrophication but rather to an increase in the temporal variation of productivity in combination with a decrease in species asynchrony in diverse communities. Our results demonstrate separate and synergistic effects of diversity and eutrophication on stability, emphasizing the need to understand how drivers of global change interactively affect the reliable provisioning of ecosystem services in real-world systems.**

Rapid declines in plant diversity have prompted concern over the consequences for the stability of ecosystem functioning and the reliable provisioning of ecological services<sup>7,12,13</sup>. The first attempts to address this concern used observational studies of natural variation in diversity<sup>14,15</sup> and were limited in their ability to separate effects of diversity from other confounding factors<sup>16</sup>. In response, more recent studies have directly manipulated diversity in experimentally established communities (that is, biodiversity experiments) to assess its impact on ecosystem functioning, particularly above-ground net primary production (ANPP)<sup>1–7</sup>. Numerous biodiversity experiments have shown that greater species diversity promotes a greater stability of productivity over time<sup>2,3,6</sup>, with asynchronous response of species to environmental fluctuations as an important underlying mechanism<sup>1,2</sup>. There would be no stabilizing

effect if species fluctuated in perfect synchrony. However, asynchrony in species response to environmental fluctuations causes declines in the biomass of some species to be compensated for by increases in others, thus buffering temporal fluctuation in the productivity of the whole community. Species asynchrony can increase the stability of aggregate functions in species-rich communities, because compensatory effects are more likely to occur when the species pool is larger and more diverse<sup>17–19</sup>.

Biodiversity experiments, in turn, have their own limitations, and their relevance to natural grassland ecosystems is debated<sup>11,12,16</sup>. For example, experimental gradients of diversity are usually assembled randomly from a local species pool, whereas in natural systems composition and diversity are influenced by a variety of factors including nutrient availability, climatic conditions and anthropogenic land use<sup>9–11</sup>. It is also likely that diversity is not the only, or even the primary, driver of the stability of ANPP<sup>20,21</sup>; however, few experiments have simultaneously addressed changes in both biodiversity and other aspects of global change.

In natural grasslands the situation is likely to be complex, because anthropogenic impacts such as climate change and eutrophication are likely to change diversity—with potential consequences for stability—as well as having their own direct effects on stability<sup>4,13,22,23</sup>. In particular, anthropogenic increases in nutrient inputs into grasslands (through direct organic and inorganic fertilization and atmospheric deposition) affect the structure and functioning of natural ecosystems worldwide<sup>8,24,25</sup>. For instance, nutrient enrichment usually increases productivity and reduces plant diversity<sup>24,25</sup>. However, the effect of eutrophication on the stability of productivity in natural grasslands remains unclear. On the basis of theory and results limited to single-site experiments<sup>22,23</sup>, we expect eutrophication to reduce the stability of productivity, because the well-known negative effects of nutrient enrichment on diversity<sup>24,25</sup> could in turn reduce species asynchrony and stability<sup>1,6,14,22,26</sup>. However, eutrophication may have additional impacts on stability that are independent of any changes in diversity. The temporal stability of ANPP is the ratio of the temporal mean to the temporal standard deviation, so an increase in stability can result from an increase in the mean, a decrease

<sup>1</sup>Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA. <sup>2</sup>Institute of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland. <sup>3</sup>Department of Wildland Resources and the Ecology Center, Utah State University, Logan, Utah 84322, USA. <sup>4</sup>Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA. <sup>5</sup>Institute for Chemistry and Biology of the Marine Environment, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany. <sup>6</sup>Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada. <sup>7</sup>Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. <sup>8</sup>School of Environmental and Forest Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>9</sup>Australian Research Council Centre of Excellence for Environmental Decisions, School of Biological Sciences, The University of Queensland, Queensland 4072, Australia. <sup>10</sup>School of Natural Sciences, Department of Zoology, Trinity College Dublin, Dublin 2, Ireland. <sup>11</sup>State Key Laboratory of Grassland and Agro-Ecosystems, Research Station of Alpine Meadow and Wetland Ecosystems, School of Life Sciences, Lanzhou University, Lanzhou 730000, China. <sup>12</sup>Department of Biology, MSC03-2020, University of New Mexico, Albuquerque, New Mexico 87131, USA. <sup>13</sup>Instituto de Investigaciones Marinas y Costeras (IIMyC) (CONICET-UNMdP), Mar del Plata 7600, Argentina. <sup>14</sup>Department of Zoology, University of Wisconsin, Madison, Wisconsin 53706, USA. <sup>15</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA. <sup>16</sup>United States Department of Agriculture Agricultural Research Service, Grassland Soil and Water Research Lab, Temple, Texas 76502, USA. <sup>17</sup>Queensland University of Technology, School of Biological Sciences, Brisbane 4000, Australia. <sup>18</sup>Department of Entomology, University of Maryland, College Park, Maryland 20742, USA. <sup>19</sup>United States Department of Agriculture Agricultural Research Service, Agroecosystem Management Research Unit, Lincoln, Nebraska 68583, USA. <sup>20</sup>Department of Forest, Rangeland and Watershed Stewardship, Colorado State University, Fort Collins, Colorado 80523, USA. <sup>21</sup>School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588, USA. <sup>22</sup>Berkeley Initiative for Global Change Biology, University of California, Berkeley, California 94720, USA. <sup>23</sup>Yunnan Academy of Biodiversity, Southwest Forestry University, Kunming 650224, China. <sup>24</sup>Department of Plant and Soil Sciences, University of Kentucky, Lexington, Kentucky 40546, USA. <sup>25</sup>Australian Research Centre for Urban Ecology, Melbourne, c/o School of Botany, University of Melbourne, Victoria 3010, Australia. <sup>26</sup>School of Biological Sciences, Monash University, Victoria 3800, Australia. <sup>27</sup>Department of Zoology, Oregon State University, Corvallis, Oregon 97331, USA. <sup>28</sup>CSIRO Ecosystem Sciences, Wembley, WA 6913, Australia. <sup>29</sup>Swiss Federal Institute for Forest, Snow and Landscape Research, 8903 Birmensdorf, Switzerland. <sup>30</sup>School of Biology, University of Leeds, Leeds LS2 9JT, UK. <sup>31</sup>National Centre for Biological Sciences, GKVK Campus, Bangalore 560065, India. <sup>32</sup>Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK.

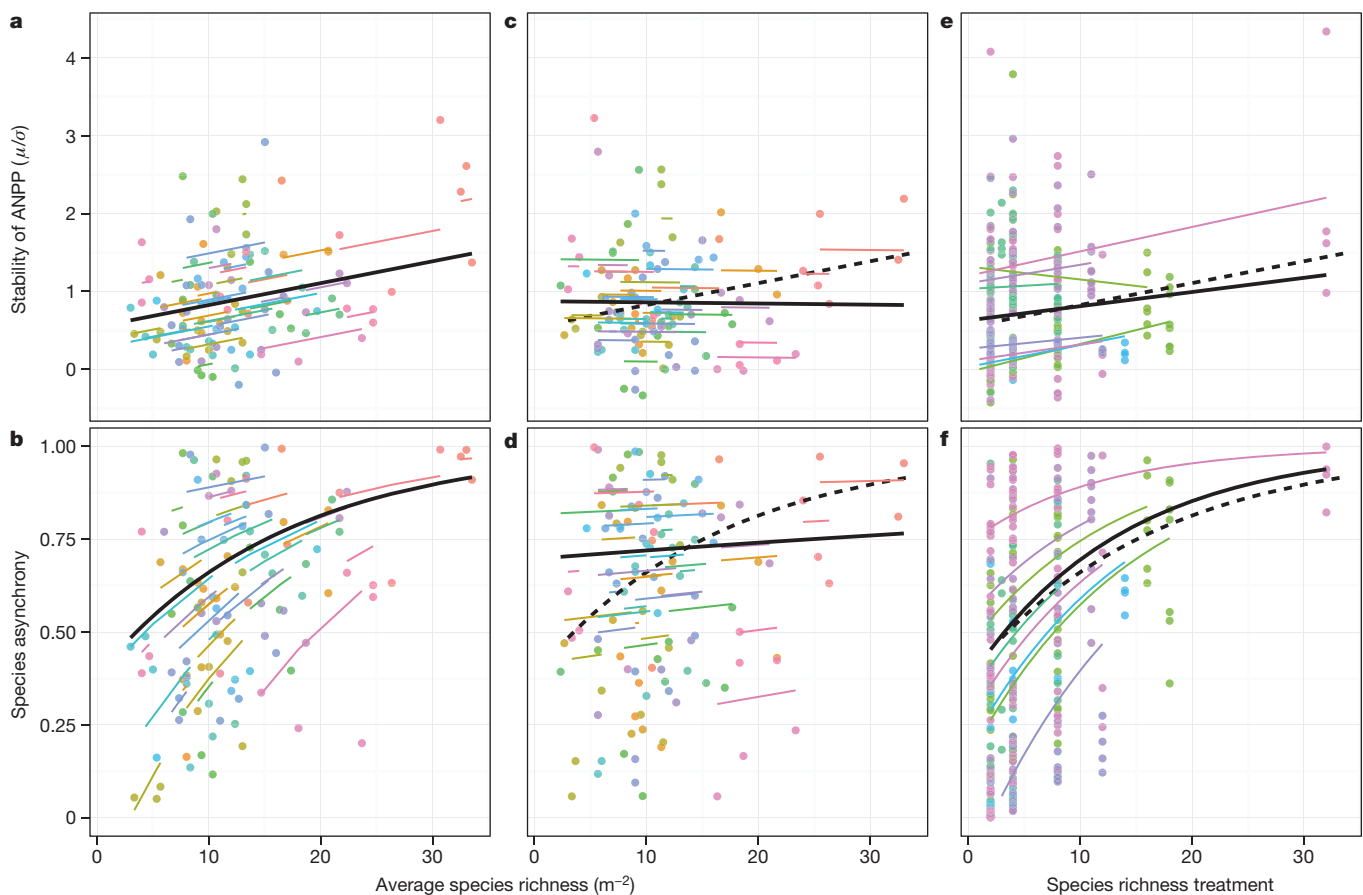
in the standard deviation, or both. Because eutrophication is expected to increase productivity it may have a stabilizing effect by increasing the temporal mean. However, there is also the potential for effects of eutrophication on stability through changes in the temporal standard deviation, but these are less well understood. We therefore require a better picture of how drivers of global change affect ecosystem stability both through changes in diversity and through other routes. Here we compare the relationship between diversity and stability found in grassland biodiversity experiments with those in fertilized and unfertilized plots in natural grasslands. We also assess the effects of eutrophication on the diversity–stability relationship both through changes in diversity and through other routes.

We evaluated the relationships between species diversity, species asynchrony and stability of ANPP across 41 naturally assembled grassland ecosystems on five continents (Extended Data Fig. 1 and Extended Data Table 1), using data from the Nutrient Network (NutNet; <http://www.nutnet.org>) collaborative experiment<sup>27,28</sup>. We used standardized methods to assess plant diversity and ANPP at each site in both unmanipulated controls and experimentally fertilized plots in a well-replicated design. We quantified diversity as the average plant species richness in standard 1-m<sup>2</sup> plots over a three-year period. Stability can take a variety of meanings in the ecological literature<sup>29,30</sup>; here we focus on temporal stability of community-level, above-ground live plant biomass from all species in a plot (a measure of ANPP) over three years. We define

temporal stability for each plot as the temporal mean of ANPP divided by its temporal variability—that is, the temporal standard deviation over a common period (see Methods).

Stability of ANPP was positively associated with plant diversity in the unmanipulated communities (Fig. 1a). Using a hierarchical sampling design and statistical model we found that stability increased with diversity consistently within and among sites, resulting in parallel relationships (coloured and black lines, respectively, in Fig. 1a). The consistent relationship between diversity and stability is concordant with experimental results obtained in grasslands across Europe<sup>1</sup> and with experiments and observations at single locations<sup>2,3,6,21,26</sup>. We used multiple regression to evaluate the influence of plant diversity and key biotic and abiotic factors on stability in our 41 grasslands. Stability was still associated with diversity after using covariates to control for differences in average site productivity and climatic conditions including annual trends, seasonality and extreme or limiting environmental factors (Extended Data Tables 1 and 2). Together these results demonstrate that temporal stability of ANPP was positively related to variation in plant diversity in our 41 naturally assembled grassland ecosystems.

We determined the role of species asynchrony as a mechanism promoting stability, by using a community-wide measure that allowed direct comparison between communities with different numbers of species<sup>17–19</sup>. Because the biomass of individual plant species was available at few sites, we used estimates based on our three-year record of the percentage



**Figure 1 | Relationships of temporal stability of ANPP (upper row) and species asynchrony (lower row) with species diversity.** **a–d**, Unmanipulated (**a**, **b**) and fertilized (**c**, **d**) communities of the Nutrient Network. **e**, **f**, The BIODEPTH network of grassland biodiversity experiments. Relationships of temporal stability of ANPP (temporal mean/temporal standard deviation; natural log transformed for analysis) of 41 grassland sites of the Nutrient Network were positive in the unmanipulated communities (**a**, **b**) (slopes and 95% confidence intervals: 0.028 (0.006 to 0.050) and 0.060 (0.023 to 0.097)), but not detectable in the fertilized communities (**c**, **d**) (−0.001 (−0.025 to 0.022) and 0.008 (−0.031 to 0.047)). (**e**, **f**) Relationships in the BIODEPTH network

were positive (0.018 (0.003 to 0.039) and 0.073 (0.053 to 0.093)). Species asynchrony varied from zero (perfect synchrony) to one (perfect asynchrony). Species richness values for the Nutrient Network are average values over the three years of post-treatment data. Points are values for individual plots ( $n = 117$  for Nutrient Network,  $n = 480$  for BIODEPTH). Black lines are the back-transformed fixed-effect linear regression slopes between sites from the mixed-effects model; coloured lines show patterns within sites. Dashed lines show regression slopes between sites in the unmanipulated communities of the Nutrient Network. Colours correspond to the ‘colour code’ column in Extended Data Table 1.

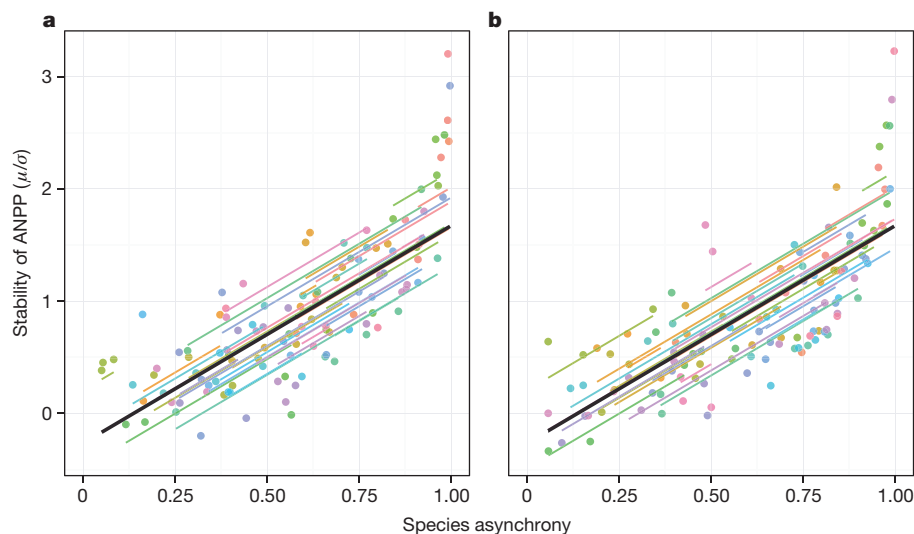
cover of individual plant species in each plot (see Methods). Our analysis of potential stability mechanisms showed that species asynchrony was positively related to plant diversity (Fig. 1b) and stability (Fig. 2a), consistent with theory on the stabilizing effects of species asynchrony in species-rich communities. Greater stability at higher diversity can also result from an increase in the temporal mean of ANPP with diversity (a 'performance-enhancing effect' that results in a higher ratio of the temporal mean to the temporal variation)<sup>1,17</sup>. Consistent with earlier NutNet analyses<sup>27</sup>, we found that the temporal mean of ANPP was not related to plant diversity (Fig. 3a). Although it is an indirect test, our result provides no support for a performance-enhancing effect in stabilizing higher-diversity communities in our study. Instead, we found stronger support for a decrease in temporal variation of ANPP (measured by the standard deviation) with diversity (Fig. 3a). In other words, greater stability at higher diversity resulted because diversity decreased the temporal variation of ANPP relative to its mean, resulting in a more stable mean-to-variance ratio.

To compare our results from naturally assembled grasslands with results from artificially assembled biodiversity experiments, we calculated values of species asynchrony from the BIODEPTH experiment<sup>1</sup>. BIODEPTH—a pan-European network of grassland biodiversity experiments—was conducted at eight field sites with a comparable hierarchical design, plot size and measurements (see Methods). Our results are comparable to those from BIODEPTH because both studies use the same three-year experimental duration and cover a similar range of diversity levels (although, by design, biodiversity experiments feature many more low-diversity communities than observational surveys). We found that the sign and slope of the overall relationships between diversity and stability and between diversity and asynchrony from our global multisite study were comparable to those from the BIODEPTH network of grassland biodiversity experiments (Fig. 1e, f; compare the solid and dashed lines).

We tested the impact of eutrophication on temporal stability and species asynchrony by using data from NutNet plots that were fertilized for three years with a combination of nitrogen, phosphorus, potassium and micronutrients (see Methods). Fertilization weakened the positive effect of diversity on stability and species asynchrony (Fig. 1c, d; compare the solid and dashed lines). We expected this result on the basis of theory, because nutrient enrichment often reduces diversity<sup>24,25</sup>, which could in turn reduce species asynchrony and stability<sup>1,6,14,22,26</sup>. However, although fertilization reduced diversity by an average of 1.3 species (95% confidence interval 0.7–1.9) per site (corresponding to a reduction

of diversity from 2.0% to 16.9% relative to average levels in the control plots ranging from 4.4 to 32.3 species per square metre (Extended Data Table 1)), counter to expectations this loss of diversity did not lead to a reduction of stability through a decrease in species asynchrony (Extended Data Fig. 2). Instead, the lower slope of the diversity–stability relationship in the fertilized communities (Fig. 1c) can be explained by a combination of two factors. First, fertilization increased the temporal variation of ANPP in diverse communities in comparison with unmanipulated communities (Fig. 3a, b; compare the dashed lines). Because fertilization generally increased mean productivity in comparison with unmanipulated communities (Fig. 3a, b; compare the solid lines), this increased variation weakened the positive effect of diversity on stability in comparison with unmanipulated communities. Second, fertilization resulted in a decrease in species asynchrony in diverse communities compared with unmanipulated communities (Fig. 1d; compare the dashed and solid lines). Because fertilization did not alter the positive relationship between species asynchrony and stability (Fig. 2b), this decrease in species asynchrony resulted in decreased stability in diverse communities compared with unmanipulated communities. In total, the results of our fertilization experiment did not show the expected destabilizing effects of diversity loss. Instead, eutrophication affected stability directly through a combination of diversity-dependent effects on species asynchrony and on the temporal variation of productivity. These direct effects of eutrophication on the diversity–stability relationship could not have been predicted from studies of natural or experimental diversity gradients.

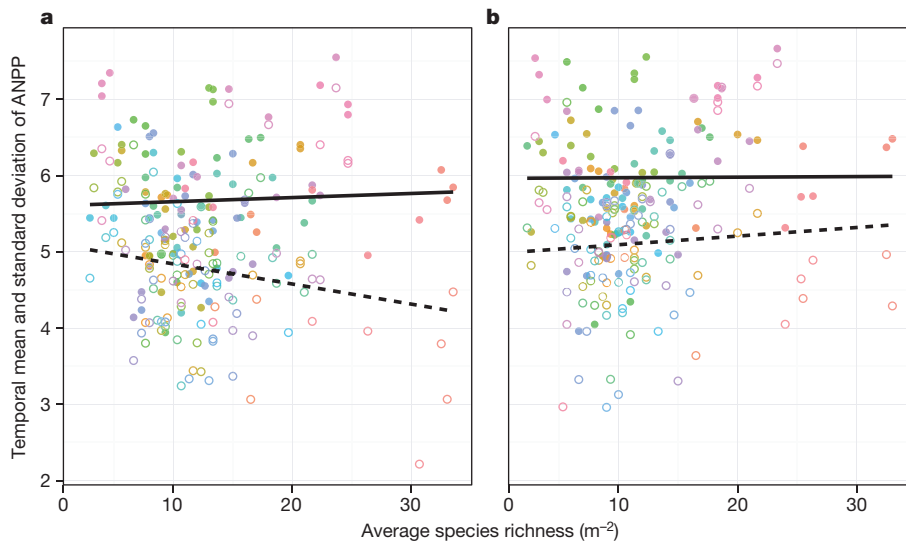
The results of our observational study of naturally assembled grassland communities are consistent with a stabilizing effect of asynchronous responses of species to environmental fluctuations in more diverse plant communities—a result previously restricted to biodiversity experiments<sup>1</sup> and observational studies at single locations<sup>21,26</sup>. However, the drivers of global change causing a loss of diversity may have additional effects on stability. The results of our fertilization experiment demonstrate impacts on stability that were not caused by changes in diversity but came about through effects of eutrophication on both the temporal variation in production and on species asynchrony. However, although the effects of fertilization on stability were not caused by species loss, the changes in species asynchrony and temporal variation that were responsible were both affected by levels of community diversity. Predicting the effects of drivers of global change therefore requires a better understanding of both their direct effects on ecosystem stability and their indirect effects through changes in diversity. Our results indicate that although



**Figure 2 | Relationships between temporal stability of ANPP (natural log transformed) and species asynchrony in 41 grassland sites of the Nutrient Network.** **a**, Unmanipulated communities; **b**, fertilized communities. The temporal stability was greatest in plots in which species fluctuations were asynchronous in both the unmanipulated (slope and 95% confidence intervals:

1.93 (1.70 to 2.16)) and fertilized communities of 41 grassland sites of the Nutrient Network (1.90 (1.58 to 2.21)). Points are values for individual plots ( $n = 117$ ). Colours correspond to the 'colour code' column in Extended Data Table 1.





**Figure 3 | Relationships of temporal mean and standard deviation of ANPP (natural log transformed) to species diversity in 41 grassland sites of the Nutrient Network. a, Unmanipulated communities; b, fertilized communities.** Temporal mean ( $\mu$ ; solid line and filled symbols) was not related to species diversity (slope and 95% confidence intervals: 0.01 (−0.02 to 0.03)) in the unmanipulated communities (a), and standard deviation ( $\sigma$ ; dashed line and open symbols) was negatively related to species diversity (−0.03 (−0.05 to −0.01)), suggesting that greater stability at higher diversity in the

eutrophication is intended to increase average levels of productivity it can also affect its temporal stability. Sustainable management of grassland ecosystems therefore requires a better understanding of the complex interrelationships between diversity, productivity and stability and how they are affected by fertilization.

## METHODS SUMMARY

The 41 sites are part of the Nutrient Network Global Research Cooperative<sup>28</sup> (Extended Data Fig. 1 and Extended Data Table 1) (see NutNet's website). Experimental plots included untreated controls and plots with nitrogen, phosphorus and potassium and micronutrients added in combination (NPK). The analyses presented here include all sites with the first three years of post-treatment community-level ANPP ( $\text{g m}^{-2} \text{ yr}^{-1}$ ) and species-level ANPP estimates based on percentage cover.

We also examined data from BIODEPTH, a consortium of coordinated biodiversity experiments that manipulated plant diversity at eight European grassland sites<sup>8</sup>. We analysed community and species-level ANPP for the three main years of this project<sup>8</sup>.

Ecosystem temporal stability was defined for each plot as  $\mu/\sigma$ , where  $\mu$  is the temporal mean of ecosystem-level ANPP and  $\sigma$  its temporal standard deviation over the three-year period. Species asynchrony was measured for each plot as  $1 - \phi_b$ , where  $\phi_b$  is species synchrony and is calculated as  $1 - \frac{\left(\sum_{i=1}^S \sigma_i\right)^2}{\sum_{i=1}^S \sigma_i^2}$ ,

where  $\sigma_i$  is the temporal standard deviation of species  $i$  in a plot with  $S$  species over the three years<sup>18</sup>. Thus, stability and species asynchrony are related such that higher levels of species asynchrony are associated with greater stability of the community as a whole<sup>14</sup>.

We modelled relationships with linear mixed-effects models by using the lme function from the nlme library in R 2.15.1. To improve normality, the ecosystem temporal stability and community-wide species asynchrony were log-transformed before analyses. Sites and blocks nested within sites for the NutNet data and sites and species composition nested within sites for BIODEPTH were treated as random effects, allowing both the intercepts and slopes of regression versus diversity to vary between sites if supported by model selection.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 October 2013; accepted 10 January 2014.

Published online 16 February 2014.

1. Hector, A. *et al.* General stabilizing effects of plant diversity on grassland productivity through population asynchrony and overyielding. *Ecology* **91**, 2213–2220 (2010).

unmanipulated communities (Fig. 1a) resulted from a decrease in temporal variation. Both temporal mean (slope and 95% confidence intervals 0.01 (−0.02 to 0.03)) and standard deviation (slope and 95% confidence intervals 0.01 (−0.02 to 0.04)) were not related to species diversity in the fertilized communities (b). Fertilization increased the temporal variation in diverse communities compared with unmanipulated communities, resulting in a reduced positive effect of diversity on stability in fertilized communities (Fig. 1c). Points are values for individual plots ( $n = 117$ ).

2. Isbell, F. I., Polley, H. W. & Wilsey, B. J. Biodiversity, productivity and the temporal stability of productivity: patterns and processes. *Ecol. Lett.* **12**, 443–451 (2009).
3. Tilman, D., Reich, P. B. & Knops, J. M. H. Biodiversity and ecosystem stability in a decade-long grassland experiment. *Nature* **441**, 629–632 (2006).
4. Reich, P. B. Elevated  $\text{CO}_2$  reduces losses of plant diversity caused by nitrogen deposition. *Science* **326**, 1399–1402 (2009).
5. Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H. & Woodfin, R. M. Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737 (1994).
6. Bezemer, T. M. & van der Putten, W. H. Ecology: diversity and stability in plant communities. *Nature* **446**, E6–E7 (2007).
7. Cardinale, B. J. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
8. Vitousek, P. M., Mooney, H. A., Lubchenco, J. & Melillo, J. M. Human domination of Earth's ecosystems. *Science* **277**, 494–499 (1997).
9. Selimant, P. C., Zavaleta, E. S., Pasari, J. R. & Hernandez, D. L. Realistic plant species losses reduce invasion resistance in a California serpentine grassland. *J. Ecol.* **100**, 723–731 (2012).
10. Zavaleta, E. S. & Hulvey, K. B. Realistic species losses disproportionately reduce grassland resistance to biological invaders. *Science* **306**, 1175–1177 (2004).
11. Srivastava, D. S. & Vellend, M. Biodiversity–ecosystem function research: is it relevant to conservation? *Annu. Rev. Ecol. Syst.* **36**, 267–294 (2005).
12. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808 (2001).
13. Ives, A. R. & Carpenter, S. R. Stability and diversity of ecosystems. *Science* **317**, 58–62 (2007).
14. Tilman, D. & Downing, J. A. Biodiversity and stability in grasslands. *Nature* **367**, 165–175 (1994).
15. McNaughton, S. J. Stability and diversity of ecological communities. *Nature* **274**, 251–253 (1978).
16. Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia* **110**, 449–460 (1997).
17. Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
18. Loreau, M. & de Mazancourt, C. Species synchrony and its drivers: neutral and nonneutral community dynamics in fluctuating environments. *Am. Nat.* **172**, E48–E66 (2008).
19. Loreau, M. *From Populations to Ecosystems: Theoretical Foundations for a New Ecological Synthesis* (Princeton Univ. Press, 2010).
20. Sankaran, M. & McNaughton, S. J. Determinants of biodiversity regulate compositional stability of communities. *Nature* **401**, 691–693 (1999).
21. Bai, Y., Han, X., Wu, J., Chen, Z. & Li, L. Ecosystem stability and compensatory effects in the inner Mongolia grassland. *Nature* **431**, 181–184 (2004).
22. Yang, Z. L., van Ruijven, J. & Du, G. Z. The effects of long-term fertilization on the temporal stability of alpine meadow communities. *Plant Soil* **345**, 315–324 (2011).
23. Yang, H. J. *et al.* Diversity-dependent stability under mowing and nutrient addition: evidence from a 7-year grassland experiment. *Ecol. Lett.* **15**, 619–626 (2012).

24. Stevens, C. J., Dise, N. B., Mountford, J. O. & Gowing, D. J. Impact of nitrogen deposition on the species richness of grasslands. *Science* **303**, 1876–1879 (2004).
25. Hautier, Y., Niklaus, P. A. & Hector, A. Competition for light causes plant biodiversity loss after eutrophication. *Science* **324**, 636–638 (2009).
26. Tilman, D. Biodiversity: population versus ecosystem stability. *Ecology* **77**, 350–353 (1996).
27. Adler, P. B. *et al.* Productivity is a poor predictor of plant species richness. *Science* **333**, 1750–1753 (2011).
28. Borer, E. T. *et al.* Finding generality in ecology: a model for globally distributed experiments. *Methods Ecol. Evol.* **5**, 65–73 (2014).
29. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, 1973).
30. Pimm, S. L. The complexity and stability of ecosystems. *Nature* **307**, 669–674 (1984).

**Acknowledgements** The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 298935 to Y.H. (with A.H. and E.W.S.). This work was generated using data from the Nutrient Network (<http://www.nutnet.org>) experiment, funded at the site-scale by individual researchers. Coordination and data management have been

supported by funding to E.T.B. and E.W.S. from the National Science Foundation Research Coordination Network (NSF-DEB-1042132), the Long Term Ecological Research (LTER) programme (NSF-DEB-1234162 to Cedar Creek as well as other LTER sites), and the Institute on the Environment at the University of Minnesota (DG-0001-13). We also thank the Minnesota Supercomputer Institute for hosting project data, and the Institute on the Environment for hosting Network meetings. We thank R. S. L. Veiga, F. Isbell, R. K. Didham, H. McGinness and M. O'Brien for suggestions that improved the manuscript.

**Author Contributions** E.W.S., E.T.B., W.S.H. and E.M.L. are Nutrient Network coordinators. Y.H. and A.H. developed and framed research questions. Y.H., E.W.S., E.T.B., P.B.A., W.S.H., H.H., A.S.M.D., C.J.S., J.D.B., Y.M.B., C.C., S.L.C., E.I.D., K.F.D., P.A.F., J.F., D.S.G., V.L.J., J.A.K., J.M.H.K., K.J.L., W.L., R.L.McC., B.A.M., J.L.M., S.M.P., A.C.R., M.S., M.S. and A.H. collected the data used in this analysis. Y.H. and A.H. analysed the data. E.W.S., E.T.B., H.H., E.M.L., P.D., K.J.L., J.L.M., L.R.O. and M.S. contributed to data analyses. Y.H. and A.H. wrote the paper with input from all authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.H. ([hauti001@umn.edu](mailto:hauti001@umn.edu)).

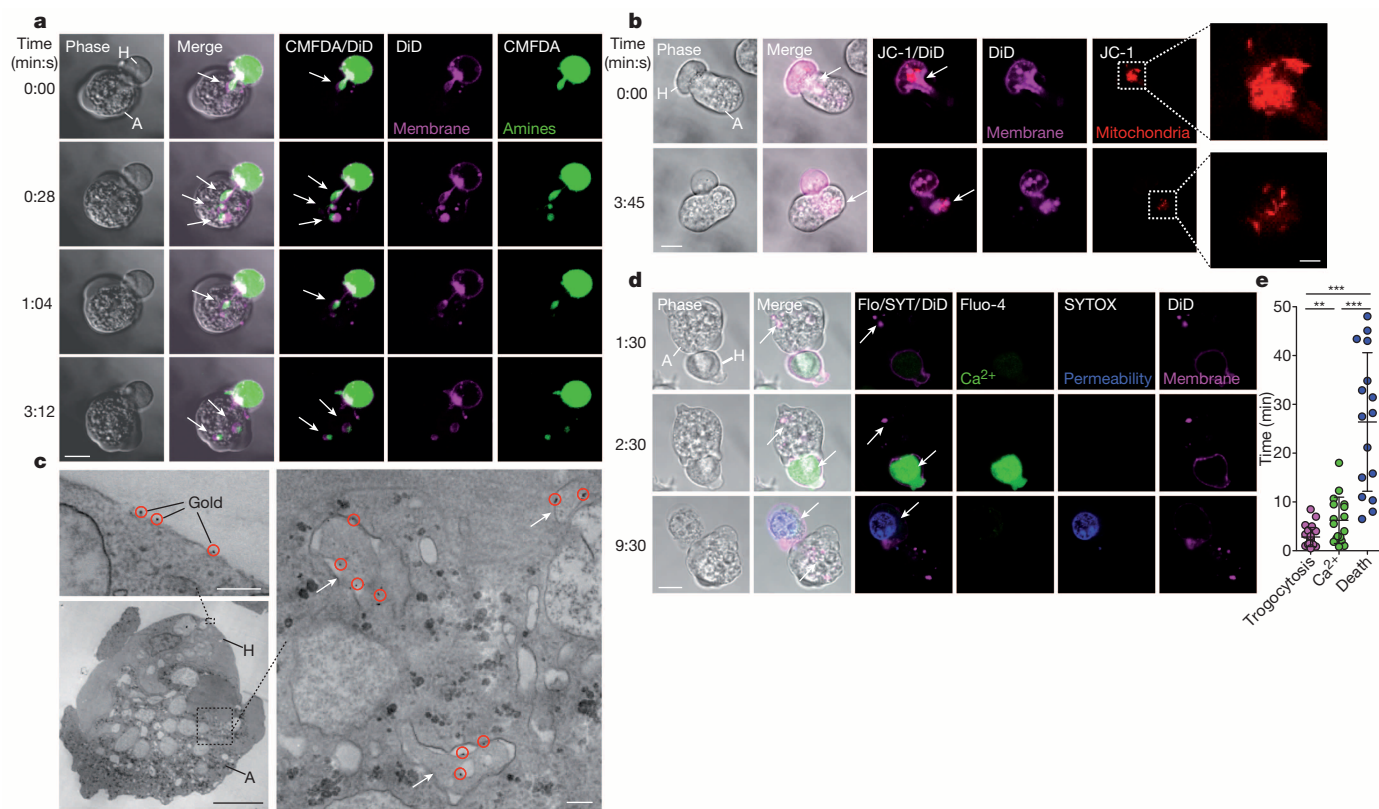
# Trogocytosis by *Entamoeba histolytica* contributes to cell killing and tissue invasion

Katherine S. Ralston<sup>1</sup>, Michael D. Solga<sup>2</sup>, Nicole M. Mackey-Lawrence<sup>1</sup>, Somlata<sup>3</sup>, Alok Bhattacharya<sup>3</sup> & William A. Petri Jr<sup>1,2,4</sup>

*Entamoeba histolytica* is the causative agent of amoebiasis, a potentially fatal diarrhoeal disease in the developing world. The parasite was named “*histolytica*” for its ability to destroy host tissues, which is probably driven by direct killing of human cells. The mechanism of human cell killing has been unclear, although the accepted model was that the parasites use secreted toxic effectors to kill cells before ingestion<sup>1</sup>. Here we report the discovery that amoebae kill by ingesting distinct pieces of living human cells, resulting in intracellular calcium elevation and eventual cell death. After cell killing, amoebae detach and cease ingestion. Ingestion of human cell fragments is required for cell killing, and also contributes to invasion of intestinal tissue. The internalization of fragments of living human cells is reminiscent of trogocytosis (from Greek *trogo*, nibble) observed between

immune cells<sup>2–6</sup>, but amoebic trogocytosis differs because it results in death. The ingestion of live cell material and the rejection of corpses illuminate a stark contrast to the established model of dead cell clearance in multicellular organisms<sup>7</sup>. These findings change the model for tissue destruction in amoebiasis and suggest an ancient origin of trogocytosis as a form of intercellular exchange.

Amoebiasis is prevalent in the developing world, and in the first year of life, approximately one third of infants in an urban slum in Dhaka, Bangladesh are infected with *E. histolytica*<sup>8</sup>. Malnourishment and stunting are associated with repeated infections in children<sup>9</sup>. During infection, amoebae colonize the colon, which can be asymptomatic or result in diarrhoea, colitis or extraintestinal disease. Invasive disease is characterized by profound tissue destruction, manifesting as massive intestinal



**Figure 1 | Amoebae internalize human cell fragments, preceding human cell death.** **a**, DiD and CMFDA-labelled human Jurkat cells (H); an amoeba (A) internalizes human cell fragments (arrows) over time. Images are representative of three independent experiments. **b**, DiD and JC-1-labelled human Jurkat cells; mitochondria (arrow) are ingested by the amoeba in an ingested fragment. Images are representative of six independent experiments. **c**, Electron microscopy with gold-labelled human Jurkat cells. Inset (top left), gold-labelled human cell membrane (gold, circles). Inset (right), fragments (arrows) within the amoeba. Images are representative of two independent

experiments. **d**, DiD and Fluo-4-labelled human Jurkat cells; with SYTOX blue present during imaging. Arrows, amoebic trogocytosis (1:30), intracellular calcium concentration elevation (2:30) and membrane permeabilization (9:30). Images are representative of 15 independent experiments. **e**, Timing in 60 cells from 15 independent experiments; shown are data points, means and standard deviations; *P* values from *t*-tests: \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001. Scale bars, 10 μm (**a**, **b**, **d**), 2 μm (**b**, insets), 0.5 μm (**c**, top left, right) and 5 μm (**c**, bottom left).

<sup>1</sup>Department of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>2</sup>Department of Microbiology, Immunology and Cancer Biology, University of Virginia, Charlottesville, Virginia 22908, USA. <sup>3</sup>School of Life Sciences, Jawaharlal Nehru University, 110067 New Delhi, India. <sup>4</sup>Department of Pathology, University of Virginia, Charlottesville, Virginia 22908, USA.



ulceration or fatal extraintestinal abscesses. This is probably driven by the cell killing activity of the amoebae, which are potently cytotoxic and capable of killing a variety of cell types including epithelial cells, immune cells and many tissue culture lines. Contact mediated by the amoeba surface galactose- and *N*-acetyl-D-galactosamine-binding (Gal/GalNAc) lectin is critical, but the molecular mechanism by which cells are killed is unknown. The established model is that the amoebae first kill cells and then ingest dead cell corpses<sup>1,10</sup> (Extended Data Fig. 1a).

To elucidate the mechanism for human cell killing, we used live microscopy to examine host–parasite interactions directly. Surprisingly, we found that, immediately after human cell contact, the amoebae internalized

distinct fragments of human cell membrane (Fig. 1a). This was initiated within one minute of human cell contact, and amoebae continued to ingest additional fragments thereafter (Supplementary Video 1). By pre-labelling human cellular compartments, we found that ingested membrane fragments contained human cell cytoplasm 90% of the time (Fig. 1a) and mitochondria 31% of the time (Fig. 1b). Pre-labelling of human cells with biotin and streptavidin–gold before co-incubation with amoebae further demonstrated that the ingested fragments were human-derived (Fig. 1c and Extended Data Fig. 2a). There was pronounced actin polymerization within the amoebae at the site of human cell contact and ingested fragments were surrounded by polymerized actin (Extended Data

## Figure 2 | Amoebic trophocytosis is predominant and specific to live human cells.

Imaging flow cytometry with CMFDA-labelled amoebae (A, blue), DiD-labelled human Jurkat cells (H, yellow) and live/dead violet-labelled dead cells (purple). **a**, Detection of internalized human material. **b**, Measurement of fragmentation of internalized human material, gated from low to high. **c**, Example merged CMFDA, DiD and bright field images. Images are representative of 10,000 images collected for each replicate. **d**, Detection of dead cells not attached to amoebae. **e**, Example images; live/dead violet (left panels) and merged CMFDA, DiD and bright field (right panels). Images are representative of 10,000 images collected for each replicate. **f**, Detection of dead cells attached to amoebae. **g**, Example images, as in **e**. Images are representative of 10,000 images collected for each replicate. **h–k**, Means and standard deviations for biological replicates (10,000 events/each). *P* values from *t*-tests: \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

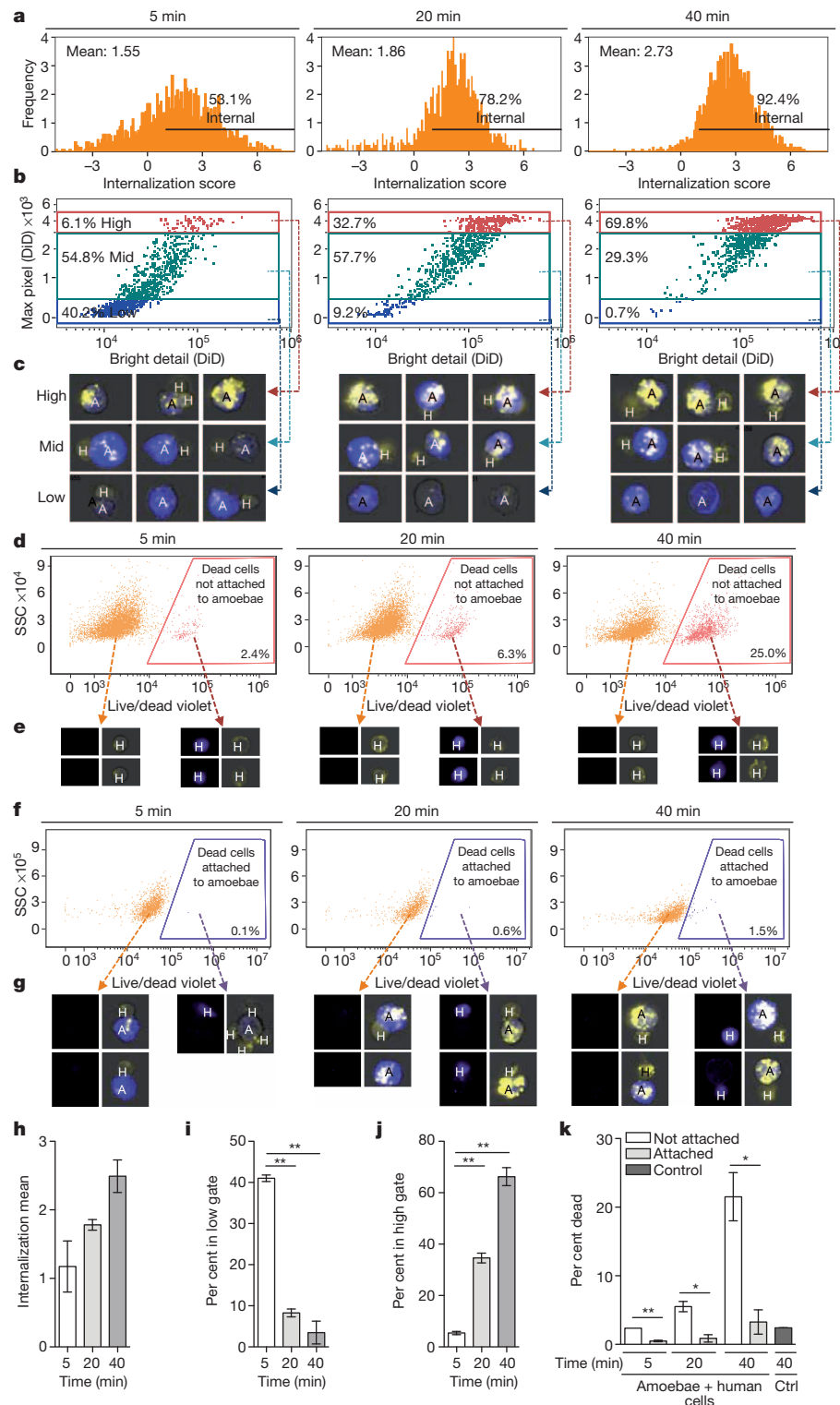


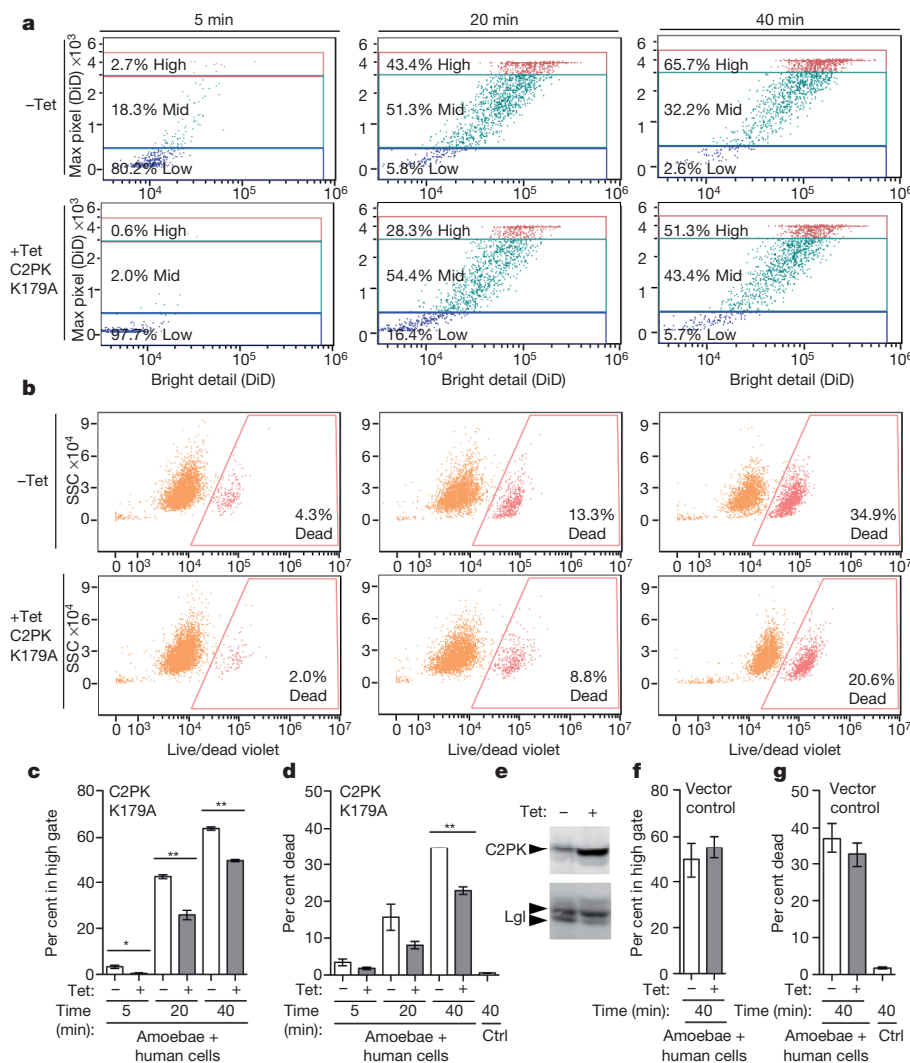
Fig. 2b, c). Ingested fragments had two membranes, with an inner human cell membrane and an outer amoeba membrane (Extended Data Fig. 2a, d).

Another unexpected finding was that human cells were alive when ingestion of fragments initiated. Irreversible intracellular calcium elevation occurred in human cells within the first few minutes, typically following detectable amoebic trophocytosis (Fig. 1d, e and Supplementary Video 2). Despite the fact that numerous fragments were physically extracted from human cells by the amoebae, human cells initially retained plasma membrane integrity (Extended Data Fig. 3a). However, human cells were eventually killed as evidenced by loss of plasma membrane integrity<sup>11</sup> (Fig. 1d, e, Extended Data Fig. 3b and Supplementary Video 1), degradation of nuclear DNA and loss of mitochondrial potential (Extended Data Fig. 4a, b). Interestingly, ingestion of fragments ceased once cells were dead (Extended Data Figs 1b and 3b and Supplementary Video 1), at which point the amoebae detached from corpses. Ingestion of fragments required living cells, because pre-killed human cells were ingested intact (Extended Data Figs 1c and 4c, d). The internalization of fragments of living cells is reminiscent of trophocytosis<sup>2–6,12</sup> (from Greek *trogo*, nibble); hence we refer to the *E. histolytica* process as ‘amoebic trophocytosis’.

To quantify amoebic trophocytosis and cell death at the population level, we used imaging flow cytometry (Fig. 2 and Extended Data Fig. 5). We found that internalization of human cellular material increased from 63.1% at 5 min to 92.4% at 40 min (Fig. 2a, h). To determine whether internalized material represented whole cells or fragments, we measured the fragmentation of the human cells (Fig. 2b, c, i, j). The population representing high fragmentation (numerous ingested human cell fragments)

increased over time (Fig. 2b, j), from 6.1% at 5 min to 69.8% at 40 min. In contrast, the population representing low fragmentation (few ingested human cell fragments), decreased over time (Fig. 2b, i). Thus, population-level quantification demonstrated that amoebic trophocytosis was predominant. The majority of dead human cells were not attached to amoebae (Fig. 2d, e, k), and very few remained attached to amoebae (Fig. 2f, g, k), supporting our observation that amoebae cease trophocytosis once cells have been killed (Extended Data Fig. 3b and Supplementary Video 1).

Neither amoebic trophocytosis nor cell killing occurred when cells were artificially brought together by centrifugation and co-incubated at 4 °C instead of 37 °C (Extended Data Figs 6a and 7f), in agreement with the known inhibition of amoebic cytotoxicity at 4 °C<sup>13</sup>. Pharmacological inhibition of amoebic phosphatidylinositol 3-kinase (PI3K) signalling using wortmannin also inhibited amoebic trophocytosis and cell death (Extended Data Fig. 6g, j). Interference with the dynamics of microfilament formation using cytochalasin D inhibited both amoebic trophocytosis and cell death (Extended Data Fig. 7a–f). Tetracycline-inducible overexpression of a kinase-dead point mutant of the amoebic C2 domain-containing protein kinase (EhC2PK)<sup>14</sup> reduced both amoebic trophocytosis and cell death (Fig. 3). EhC2PK is an early regulator of phagocytosis in *E. histolytica*<sup>14</sup>, thus some initiators of phagocytosis may be common to amoebic trophocytosis. To examine signalling upstream of EhC2PK, we used blocking antibodies to interfere with the amoebic Gal/GalNAc lectin, which has been shown to have a role in initiating cell killing<sup>1,15,16</sup> separable from its role in host-cell attachment<sup>13</sup>. An anti-Gal/GalNAc lectin blocking monoclonal antibody directed to epitope 1, which reduces



**Figure 3 | Amoebic trophocytosis contributes to human cell killing.** a–d, Imaging flow cytometry analysis with amoebae expressing a tetracycline (Tet)-inducible copy of EhC2PK K179A.

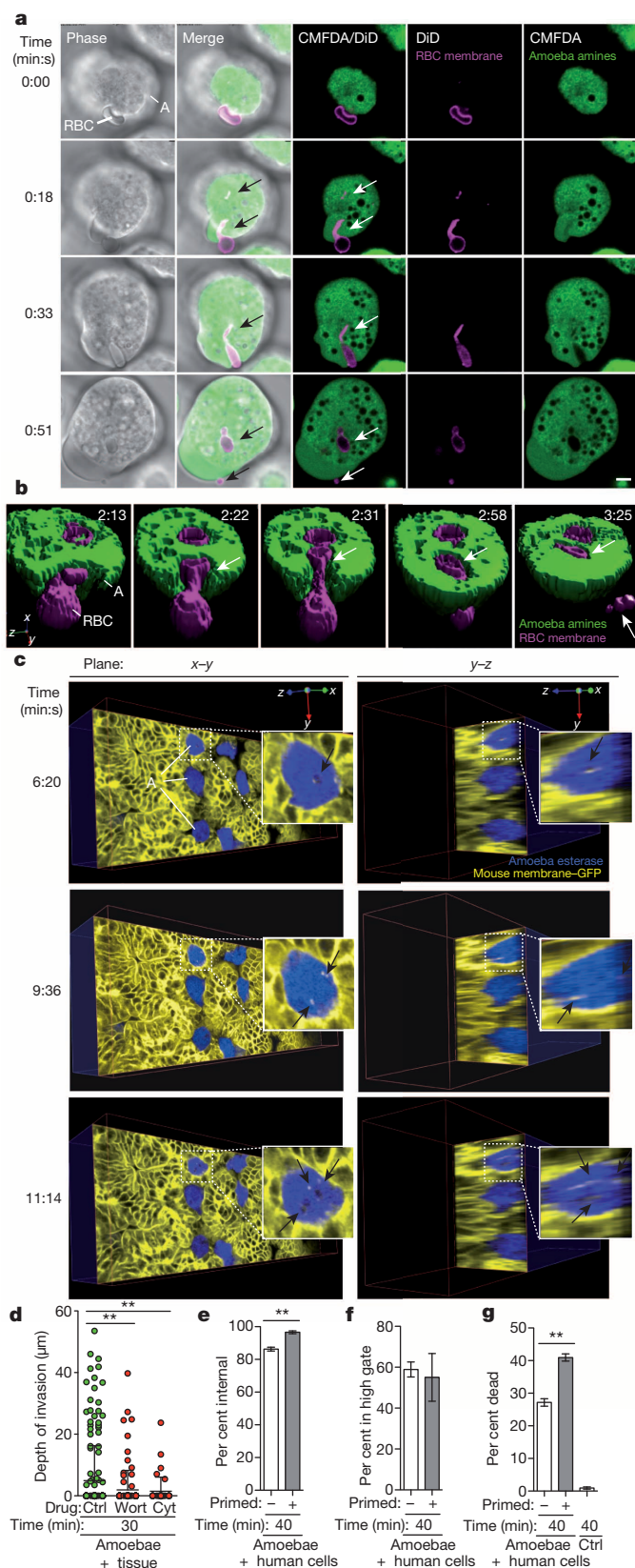
a, Measurement of fragmentation. b, Detection of dead cells. c, d, Means and standard deviations for biological duplicates at each time point (10,000 events/each). e, Western blot demonstrating EhC2PK overexpression in Tet-induced cells; Lgl, loading control. Blots are representative of three independent experiments. f, g, Imaging flow cytometry analysis with vector control amoebae; shown are means and standard deviations for biological duplicates (10,000 events/each). P values from t-tests: \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

killing while enhancing attachment<sup>16</sup>, inhibited both amoebic trogocytosis and cell death compared to a control monoclonal antibody directed to epitope 3 (Extended Data Fig. 7g–i). Thus amoebic trogocytosis contributes to human cell killing and requires physiological temperature, actin rearrangements and Gal/GalNAc lectin, EhC2PK and PI3K signalling.

The hallmark of amoebiasis is the detection of amoebae with ingested red blood cells<sup>17</sup>. Live imaging demonstrated that amoebic trogocytosis occurred during red blood cell ingestion (Fig. 4a, b and Supplementary Videos 3 and 4). We also detected amoebic trogocytosis of human colonic epithelial cells (Extended Data Fig. 8). Intriguingly, amoebic trogocytosis occurred when amoebae were incubated with *ex vivo* mouse intestinal tissue and monitored live in three dimensions using two-photon microscopy (Fig. 4c and Supplementary Videos 5 and 6). Amoebae traversed the intestinal crypts, as has been demonstrated in studies using *ex vivo* human intestine<sup>18</sup>. Tissue cell death occurred and amoebic trogocytosis of intestinal enterocytes preceded amoebic invasion into the tissue (Extended Data Fig. 9 and 10 and Supplementary Videos 7 and 8), further supporting its pathologic relevance. Moreover, inhibition of amoebic trogocytosis by treatment with wortmannin or cytochalasin D significantly blocked tissue invasion (Fig. 4d). Further supporting the relevance of amoebic trogocytosis to pathogenesis, we found that amoebae that had previously undergone amoebic trogocytosis were primed to undergo more ingestion and more cell killing (Fig. 4e–g) than amoebae that had not undergone trogocytosis. More of the primed amoebae underwent trogocytosis (Fig. 4e), although they ingested the same number of fragments as non-primed amoebae (Fig. 4f), indicating that priming enhances the initiation of amoebic trogocytosis.

Therefore several lines of evidence demonstrate that amoebic trogocytosis contributes to cell killing and is likely to contribute to pathogenesis *in vivo*. Cell death probably stems from the accumulation of membrane damage. Targeted cells initially retain membrane integrity, hence ingestion of one fragment is not enough to kill. Inhibition experiments demonstrated that a reduction in the number of ingested fragments almost completely prevented cell death, indicating that there is a threshold of tolerated damage. It has previously been reported that amoebae ingest killed cells<sup>10</sup>, but in these studies the target cells were pre-killed and probably had different characteristics from cells killed directly by the amoebae. Consistent with this idea, we found that pre-killed cells were internalized whole and thus they were recognized differently than endogenously killed cells, which the amoebae ceased ingesting.

The rejection of dead cell corpses is remarkable and makes a nutritional role for amoebic trogocytosis seem unlikely. Intestinal bacteria may be the primary food source during infection and the energetic costs of full internalization of human cells may outweigh the benefits. The large size of epithelial cells and the tight intercellular connections may make ingestion of entire cells difficult. Amoebic trogocytosis could potentially lower tissue density and create a more spacious environment for amoeboid migration. Ingesting fragments could also provide the amoebae with a means of sensing the environment. Interestingly, amoebae that had undergone amoebic trogocytosis were primed to undergo more trogocytosis, suggesting that there might be ‘feed-forward’ regulation, similar to what has been demonstrated during the ingestion of



**Figure 4 | Amoebic trogocytosis occurs with red blood cells, contributes to intestinal invasion, and promotes enhanced cell killing.** **a**, DiD-labelled red blood cells (RBC) and CMFDA-labelled amoebae (A); ingested fragments (arrows) and a fragment that remains extracellular (0:51). Scale bar, 5 μm. Images are representative of three independent experiments. **b**, Example three dimensional reconstruction with DiD-labelled red blood cells and CMFDA-labelled amoebae; ingested fragments (arrows) and a fragment that remains extracellular (3:25). Images are representative of three independent experiments. **c**, Planes from live three-dimensional two-photon microscopy with mouse intestinal tissue from a mouse expressing membrane-targeted EGFP (yellow); calcein violet-labelled amoebae (blue) are ingesting fragments (arrows). Images are representative of three independent experiments. **d**, Amoebae treated with wortmannin or cytochalasin D were incubated with tissue. Invasion depth, means and standard deviations from biological replicates (at least 40 cells per treatment); *P* values are indicated. **e–g**, Amoebae were incubated with or without human Jurkat cells (‘primed’), before incubation with labelled human Jurkat cells. **e**, Measurement of internalization. **f**, Measurement of fragmentation. **g**, Detection of dead cells. Means and standard deviations for biological duplicates (10,000 events/each). *P* values from *t*-tests: \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.



beads and apoptotic cells<sup>19</sup>. Therefore amoebic trogocytosis and destruction of cells may promote further destruction. Further work will be needed to confirm with certainty that the proposed model for amoebic trogocytosis in cell killing and tissue invasion is applicable to pathogenesis *in vivo*, and to explore the implications of this process in disease.

Distortion of target cell shape during amoebic ingestion has been reported previously<sup>20,21</sup>. We have similarly observed distortion before the appearance of fragments. We used a variety of membrane labels, and fragments are more readily detected using these tightly localized and bright labels, than with more diffuse probes. It is therefore possible that, owing to differences in detection methods and the dynamic and rapid appearance of fragments, they were not previously observed. Because increased red blood cell rigidity reduces cell distortion during ingestion<sup>22</sup>, it is also conceivable that differences in the deformability of different cell types influence the extent of fragmentation that occurs. Less deformable targets like beads are likely to be ingested whole. Finally, the size of the target cell type determines the amount of material that remains extracellular; very little remains following red blood cell ingestion, whereas much more remains following Jurkat or epithelial cell ingestion.

Some mechanistic details of trogocytosis between immune cells are beginning to be defined<sup>23</sup>. A key difference between immune cell trogocytosis and amoebic trogocytosis is that immune trogocytosis does not result in cell death. This may be because immune trogocytosis involves the exchange of fewer fragments, and we find that targeted cells can withstand the ingestion of a limited number of fragments. Interestingly, there are hints that protozoa like *Naegleria fowleri* may also be capable of ingesting host cell fragments<sup>12</sup>. Therefore, trogocytosis as a form of intercellular exchange may be more evolutionarily ancient and widespread than is currently appreciated. Finally, amoebic trogocytosis is a potentially promising target for the future development of new therapeutics for amoebiasis, a major neglected disease in the developing world.

## METHODS SUMMARY

Amoebae (HM1:IMSS), human Jurkat cells and human red blood cells were prepared as described<sup>10,24</sup>. In some experiments, parasites were stably transfected with a plasmid for tetracycline-inducible expression of EhC2PK K179A<sup>14</sup>. Mice with membrane-targeted enhanced green fluorescent protein (EGFP)<sup>25</sup> were used in compliance with the University of Virginia Institutional Animal Care and Use Committee, and PHS and USDA guidelines. For live imaging, Jurkat cells were pre-labelled with DiD, DiI, CMFDA and/or JC-1 (Invitrogen) and were combined with amoebae immediately before imaging on a Zeiss LSM 510 or LSM 700 confocal microscope. In some experiments, SYTOX blue (Invitrogen) was present during imaging. Human red blood cells were labelled with DiD. For electron microscopy, Jurkat cells were pre-labelled with Sulfo-NH-SS biotin (Thermo) and streptavidin-6 nm gold (Electron Microscopy Sciences). Amoebae and Jurkat cells were combined before fixation, processing and imaging on a JEOL 1230 microscope. For imaging flow cytometry, amoebae were pre-labelled with CMFDA and Jurkat cells with sulphonated DiD. Pharmacological interventions were carried out in the amoebae after CMFDA labelling, using wortmannin or cytochalasin D (Sigma). At each time point, samples were transferred to ice and labelled with live/dead fixable violet (Invitrogen) before fixation. Samples were collected on an Amnis ImageStreamX or ImageStreamX Mark II flow cytometer. 10,000 to 15,000 events were collected for each sample and data were analysed using Amnis IDEAS. For *ex vivo* imaging, amoebae were labelled with calcein violet (Invitrogen). Mouse caeca were isolated and amoebae added immediately before imaging with a FLUOVIEW FV1000 Olympus multiphoton system. Three dimensional reconstructions were made with Velocity (PerkinElmer). Depth of tissue invasion was quantified in a blinded fashion in ten non-consecutive frozen sections per treatment. Quantitative data are expressed as the mean and standard deviation of the mean. Statistical significance was determined with the unpaired two-tailed Student's *t*-test (Prism, GraphPad Software, Inc.).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 January 2013; accepted 10 March 2014.

Published online 9 April 2014.

1. Ralston, K. S. & Petri, W. A. Jr. Tissue destruction and invasion by *Entamoeba histolytica*. *Trends Parasitol.* **27**, 254–263 (2011).

2. Batista, F. D., Iber, D. & Neuberger, M. S. B cells acquire antigen from target cells after synapse formation. *Nature* **411**, 489–494 (2001).
3. Huang, J. F. *et al.* TCR-mediated internalization of peptide-MHC complexes acquired by T cells. *Science* **286**, 952–954 (1999).
4. Hudrisier, D., Riond, J., Mazarguil, H., Gairin, J. E. & Joly, E. Cutting edge: CTLs rapidly capture membrane fragments from target cells in a TCR signaling-dependent manner. *J. Immunol.* **166**, 3645–3649 (2001).
5. Hudson, L., Sprent, J., Miller, J. F. & Playfair, J. H. B cell-derived immunoglobulin on activated mouse T lymphocytes. *Nature* **251**, 60–62 (1974).
6. Joly, E. & Hudrisier, D. What is trogocytosis and what is its purpose? *Nature Immunol.* **4**, 815 (2003).
7. Elliott, M. R. & Ravichandran, K. S. Clearance of apoptotic cells: implications in health and disease. *J. Cell Biol.* **189**, 1059–1070 (2010).
8. Korpe, P. S. *et al.* Breast milk parasite-specific antibodies and protection from amebiasis and cryptosporidiosis in Bangladeshi infants: a prospective cohort study. *Clin. Infect. Dis.* **56**, 988–992 (2013).
9. Mondal, D., Petri, W. A. Jr, Sack, R. B., Kirkpatrick, B. D. & Haque, R. *Entamoeba histolytica*-associated diarrheal illness is negatively associated with the growth of preschool children: evidence from a prospective study. *Trans. R. Soc. Trop. Med. Hyg.* **100**, 1032–1038 (2006).
10. Huston, C. D., Boettner, D. R., Miller-Sims, V. & Petri, W. A. Jr. Apoptotic killing and phagocytosis of host cells by the parasite *Entamoeba histolytica*. *Infect. Immun.* **71**, 964–972 (2003).
11. Kroemer, G. *et al.* Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. *Cell Death Differ.* **16**, 3–11 (2009).
12. Brown, T. Observations by immunofluorescence microscopy and electron microscopy on the cytopathogenicity of *Naegleria fowleri* in mouse embryo-cell cultures. *J. Med. Microbiol.* **12**, 363–371 (1979).
13. Ravdin, J. I. & Guerrant, R. L. Role of adherence in cytopathogenic mechanisms of *Entamoeba histolytica*. Study with mammalian tissue culture cells and human erythrocytes. *J. Clin. Invest.* **68**, 1305–1313 (1981).
14. Somlata, Bhattacharya S. & Bhattacharya, A. A C2 domain protein kinase initiates phagocytosis in the protozoan parasite *Entamoeba histolytica*. *Nature Commun.* **2**, 230 (2011).
15. Ravdin, J. I., Croft, B. Y. & Guerrant, R. L. Cytopathogenic mechanisms of *Entamoeba histolytica*. *J. Exp. Med.* **152**, 377–390 (1980).
16. Saffer, L. D. & Petri, W. A. Jr. Role of the galactose lectin of *Entamoeba histolytica* in adherence-dependent killing of mammalian cells. *Infect. Immun.* **59**, 4681–4683 (1991).
17. González-Ruiz, A. *et al.* Value of microscopy in the diagnosis of dysentery associated with invasive *Entamoeba histolytica*. *J. Clin. Pathol.* **47**, 236–239 (1994).
18. Bansal, D. *et al.* An ex-vivo human intestinal model to study *Entamoeba histolytica* pathogenesis. *PLoS Negl. Trop. Dis.* **3**, e551 (2009).
19. Sateriale, A., Vaithilingam, A., Donnelly, L., Miller, P. & Huston, C. D. Feed-forward regulation of phagocytosis by *Entamoeba histolytica*. *Infect. Immun.* **80**, 4456–4462 (2012).
20. Lejeune, A. & Gicquaud, C. Evidence for two mechanisms of human erythrocyte endocytosis by *Entamoeba histolytica*-like amoebae (Laredo strain). *Biol. Cell* **59**, 239–245 (1987).
21. Nakada-Tsukui, K., Okada, H., Mitra, B. N. & Nozaki, T. Phosphatidylinositol-phosphates mediate cytoskeletal reorganization during phagocytosis via a unique modular protein consisting of RhoGEF/DH and FYVE domains in the parasitic protozoan *Entamoeba histolytica*. *Cell. Microbiol.* **11**, 1471–1491 (2009).
22. Lejeune, A. & Gicquaud, C. Target cell deformability determines the type of phagocytic mechanism used by *Entamoeba histolytica*-like, Laredo strain. *Biol. Cell* **74**, 211–216 (1992).
23. Martínez-Martin, N. *et al.* T cell receptor internalization from the immunological synapse is mediated by TC21 and RhoG GTPase-dependent phagocytosis. *Immunity* **35**, 208–222 (2011).
24. Boettner, D. R. *et al.* *Entamoeba histolytica* phagocytosis of human erythrocytes involves PATMK, a member of the transmembrane kinase family. *PLoS Pathog.* **4**, e8 (2008).
25. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. A. Redick and S. J. Guillot for assistance with sample preparation for electron microscopy and D. A. Zemo of Olympus for assistance with multiphoton microscopy. We thank the University of Virginia Research Histology Core for assistance with preparation of frozen sections. We thank J. E. Casanova, J. D. Castle, J. Lannigan, K. S. Ravichandran and R. P. Taylor for helpful discussions. The artwork (Extended Data Fig. 1) was prepared by A. Impagliazzo. K.S.R. was supported by a Howard Hughes Medical Institute Postdoctoral Fellowship from the Life Sciences Research Foundation, and a Postdoctoral Fellowship from the Hartwell Foundation. N.M.M.-L. was supported by NIH Training Grant AI07046-32. This work was supported by NIH grant 5R01 AI-26649 to W.A.P.

**Author Contributions** K.S.R. designed, performed and analysed the experiments. W.A.P. oversaw experimental design and analysis. M.D.S. assisted with Amnis ImageStream experimental design and with collection and analysis of Amnis ImageStream data. N.M.M.-L. assisted with isolation and preparation of mouse tissue for *ex vivo* imaging. S. and A.B. contributed plasmids and antibodies for the study of EhC2PK and contributed to analysis. K.S.R. and W.A.P. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.A.P. ([wap3g@virginia.edu](mailto:wap3g@virginia.edu)).

# Inhibition of *miR*-25 improves cardiac contractility in the failing heart

Christine Wahlquist<sup>1\*</sup>, Dongtak Jeong<sup>2\*</sup>, Agustin Rojas-Muñoz<sup>1</sup>, Changwon Kho<sup>2</sup>, Ahyoung Lee<sup>2</sup>, Shinichi Mitsuyama<sup>2</sup>, Alain van Mil<sup>1,3</sup>, Woo Jin Park<sup>4</sup>, Joost P. G. Sluijter<sup>3</sup>, Pieter A. F. Doevendans<sup>3</sup>, Roger J. Hajjar<sup>2</sup> & Mark Mercola<sup>1</sup>

Heart failure is characterized by a debilitating decline in cardiac function<sup>1</sup>, and recent clinical trial results indicate that improving the contractility of heart muscle cells by boosting intracellular calcium handling might be an effective therapy<sup>2,3</sup>. MicroRNAs (miRNAs) are dysregulated in heart failure<sup>4,5</sup> but whether they control contractility or constitute therapeutic targets remains speculative. Using high-throughput functional screening of the human microRNAome, here we identify miRNAs that suppress intracellular calcium handling in heart muscle by interacting with messenger RNA encoding the sarcoplasmic reticulum calcium uptake pump SERCA2a (also known as ATP2A2). Of 875 miRNAs tested, *miR*-25 potently delayed calcium uptake kinetics in cardiomyocytes *in vitro* and was upregulated in heart failure, both in mice and humans. Whereas adeno-associated virus 9 (AAV9)-mediated overexpression of *miR*-25 *in vivo* resulted in a significant loss of contractile function, injection of an antisense oligonucleotide (antagomiR) against *miR*-25 markedly halted established heart failure in a mouse model, improving cardiac function and survival relative to a control antagomiR oligonucleotide. These data reveal that increased expression of endogenous *miR*-25 contributes to declining cardiac function during heart failure and suggest that it might be targeted therapeutically to restore function.

Heart failure is the culmination of diverse cardiovascular diseases, including hypertension, ischaemic disease and atherosclerosis, valvular insufficiency, myocarditis, and contractile protein mutations, and is uniformly characterized by a progressive loss of contractile function and reserve<sup>1</sup>. The prevailing therapeutic strategy is to block the deleterious effects of the renin–angiotensin and sympathetic systems, but existing drugs target few mechanisms within the failing cardiomyocyte and there is a critical need for novel drugs, especially for treating patients with advanced heart failure<sup>6,7</sup>. A complex intracellular network balances contractility and intracellular Ca<sup>2+</sup> handling in relation to workload<sup>8</sup>; however, the role of miRNAs in cardiac contractility remains largely unexplored. miRNAs fine tune nearly all normal and pathological processes by downregulating proteins that occupy key nodal points in biological control networks<sup>9,10</sup>. We reasoned that miRNAs that repress contractility<sup>11</sup> might be upregulated during human heart failure, and might therefore constitute novel targets for therapeutic intervention.

The calcium-transporting ATPase SERCA2a is the primary mechanism of Ca<sup>2+</sup> uptake during excitation–contraction coupling in cardiomyocytes. Impaired Ca<sup>2+</sup> uptake resulting from decreased expression and reduced activity of SERCA2a is a hallmark of heart failure<sup>2</sup>. Accordingly, restoration of SERCA2a by gene transfer has proven effective in improving key parameters of heart failure in animal models<sup>12</sup> and more recently in clinical trials<sup>3</sup>. Thus, miRNAs that downregulate SERCA2a might be elevated in heart failure and compromise cardiac function. Computational algorithms predict >300 miRNAs that target SERCA2a (miRNA Data Integration Portal<sup>13</sup>), but they are error prone, necessitating

empirical evaluation<sup>14</sup>. We therefore functionally screened a whole-genome collection of miRNAs for selective downregulation of the Ca<sup>2+</sup> pump. To perform the screen, we fused the *SERCA2a* mRNA 3' untranslated region (UTR) sequence downstream of an enhanced green fluorescent protein (eGFP) coding region, making a 'target sensor' construct to permit detection of active miRNAs by a decrease in eGFP fluorescence (Fig. 1a, b). In the primary screen, 144 miRNAs reduced eGFP fluorescence by >30% with  $P < 0.05$  (Fig. 1c, d), and 82 were confirmed by testing through a dose range (Fig. 1e, f and Supplementary Table 1). Thirty-two were evolutionarily conserved and 15 were both evolutionarily conserved and reported to be upregulated in human heart failure (see Methods) (Fig. 1e). Four caused a highly significant ( $P < 0.001$ , one-tailed analysis of variance (ANOVA)) prolongation of the decay phase of the Ca<sup>2+</sup> transient in the cardiomyocyte HL-1 cell line, shown here as the duration from 75% to 25% maximal value, CaTD<sub>75–25</sub> (Fig. 1g), measured using an automated platform<sup>15</sup>. The most potent miRNA was *miR*-25, which elicited a physiological effect comparable to that of the short interfering RNA (siRNA) directed against *SERCA2a*, *siSERCA2a* (Fig. 1h). We confirmed that *miR*-25 is upregulated in human myocardial samples from patients with severe heart failure (Fig. 1i). *In situ* hybridization revealed that *miR*-25 is expressed primarily in cardiomyocytes of trans-aortic constriction (TAC)-induced failing mouse hearts, with no detectable expression in cardiac fibroblasts or vascular endothelial cells, and low expression in vascular smooth muscle (Fig. 2).

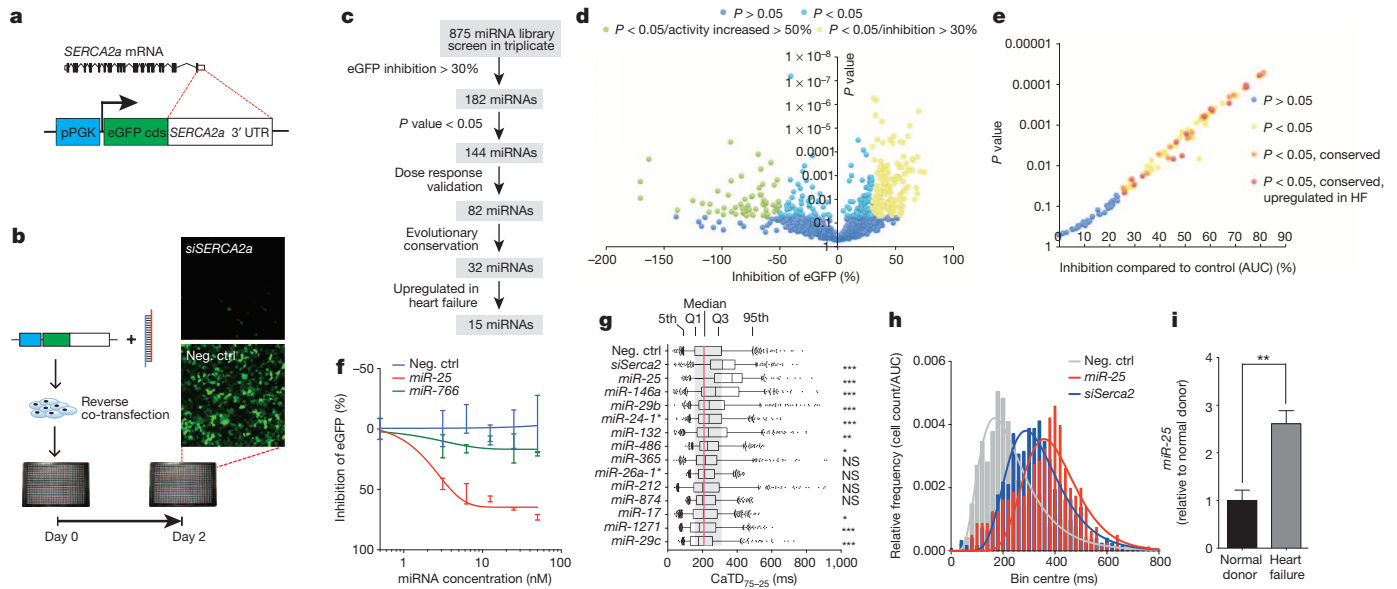
To establish a link between *miR*-25 and cardiac function, we identified putative protein targets involved in Ca<sup>2+</sup> handling computationally (DianaLab miRPath with TargetScan Mouse v.5.0, see Methods). In addition to *SERCA2a*, these included inositol-1,4,5-trisphosphate receptor 1 (IP3R1), which was selectively downregulated by *miR*-25 transient transfection in HL-1 cells (Fig. 3a and Extended Data Fig. 1a). Other candidate proteins involved in calcium handling, including the sodium–calcium exchanger (NCX1; also known as SLC8A1), calmodulin kinase 2 (CaMKII), phospholamban (PLN) and calmodulin 3 (CALM3) were unaffected in HL-1 cells transfected with *miR*-25 (Extended Data Fig. 2). miRNAs bind cognate mRNAs by imprecise base pairing to specific recognition elements<sup>9,10</sup>. Single putative *miR*-25 recognition sites were identified within *SERCA2a* and *IP3R1* 3' UTR sequences by TargetScan Human v.6.2 (Fig. 3b and Extended Data Fig. 1b). Mutation of these sequences abolished the ability of *miR*-25 to inhibit reporter expression (Fig. 3b and Extended Data Fig. 1b), further supporting selective *miR*-25 interactions with the mRNAs, and indicating that the single recognition elements identified in the 3' UTRs of the *SERCA2a* and *IP3R1* mRNAs are sufficient for *miR*-25 activity.

Having shown that *miR*-25 regulates *SERCA2a* and *IP3R1*, we evaluated whether siRNAs directed against these proteins could mimic the effect of *miR*-25 on cardiac Ca<sup>2+</sup> transient kinetics *in vitro*. Transfection of *siSerca2a* significantly slowed the decay (Ca<sup>2+</sup> re-uptake) phase of the transient (CaTD<sub>75–25</sub>) relative to controls in spontaneously contracting

<sup>1</sup>Department of Bioengineering, University of California, San Diego, and the Muscle Development and Regeneration Program, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, California 92037, USA. <sup>2</sup>The Cardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. <sup>3</sup>Department of Cardiology, University Medical Center Utrecht and ICIN Netherlands Heart Institute, Heidelberglaan 100, room G02.523, 3584 CX Utrecht, The Netherlands. <sup>4</sup>Global Research Laboratory, Gwangju Institute of Science and Technology, 123 Cheomdan-gwagiro, Buk-gu, Gwangju 500-712, South Korea.

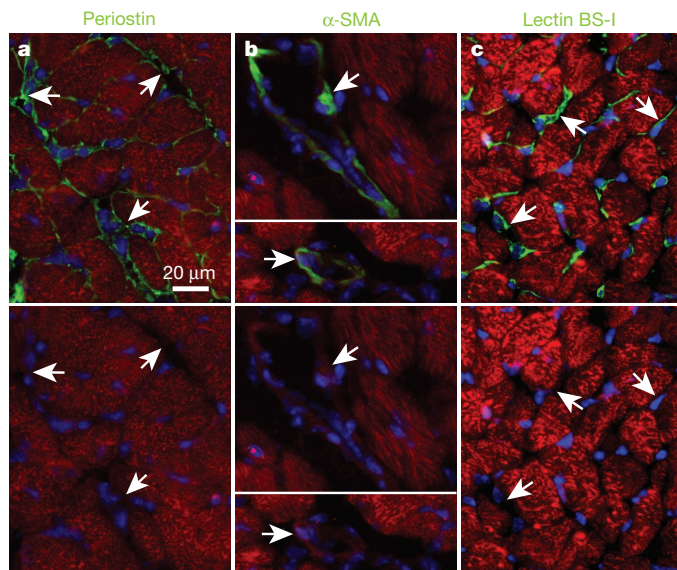
\*These authors contributed equally to this work.





**Figure 1 | High-content screening identifies miRNAs that control SERCA2a.** **a, b,** Target sensor construct (**a**) and screening workflow (**b**). cds, coding DNA sequence; Neg. ctrl, negative control; pPGK, phosphoglycerate kinase promoter. **c,** Screening summary. **d, e,** Primary ( $n = 3$ ) (**d**) and confirmatory ( $n = 3$ ) (**e**) screen data plotted as percentage inhibition relative to siRNA against *SERCA2a* (100% inhibition) and scrambled sequence control (0% inhibition) ( $x$ -axis; AUC, area under the curve) and  $P$  value from Student's  $t$ -test ( $y$ -axis). HF, heart failure. **f,** eGFP–SERCA2a inhibition by *miR-25* and inactive *miR-766* relative to scrambled sequence control ( $n = 10$ ). **g,**  $\text{Ca}^{2+}$  transient kinetic analysis of HL-1 cells transfected with miRNAs that inhibited eGFP–SERCA2a (panel **e**) and are evolutionarily conserved and upregulated in

HL-1 cells (Fig. 3c and Extended Data Fig. 1c) and neonatal rat ventricular cardiomyocytes (NRVCs; Extended Data Fig. 1d); however, the maximal velocity of the upstroke phase of the  $\text{Ca}^{2+}$  transient ( $V_{\text{max}}$ ) was unchanged (Extended Data Fig. 3), accurately reproducing the effect of



**Figure 2 | Endogenous *miR-25* expression in the heart.** **a–c,** *In situ* hybridization, revealing endogenous *miR-25* expression in failing LV myocardium (red) compared with periostin (**a**, top panel versus bottom panel),  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA) (**b**) and lectin BS-1 (**c**) staining (each in green). Hoechst 33342 staining marks nuclei (blue). Arrows indicate examples of non-cardiomyocytes. Scale bar, 20  $\mu\text{m}$ . Data are representative of two biological replicates.

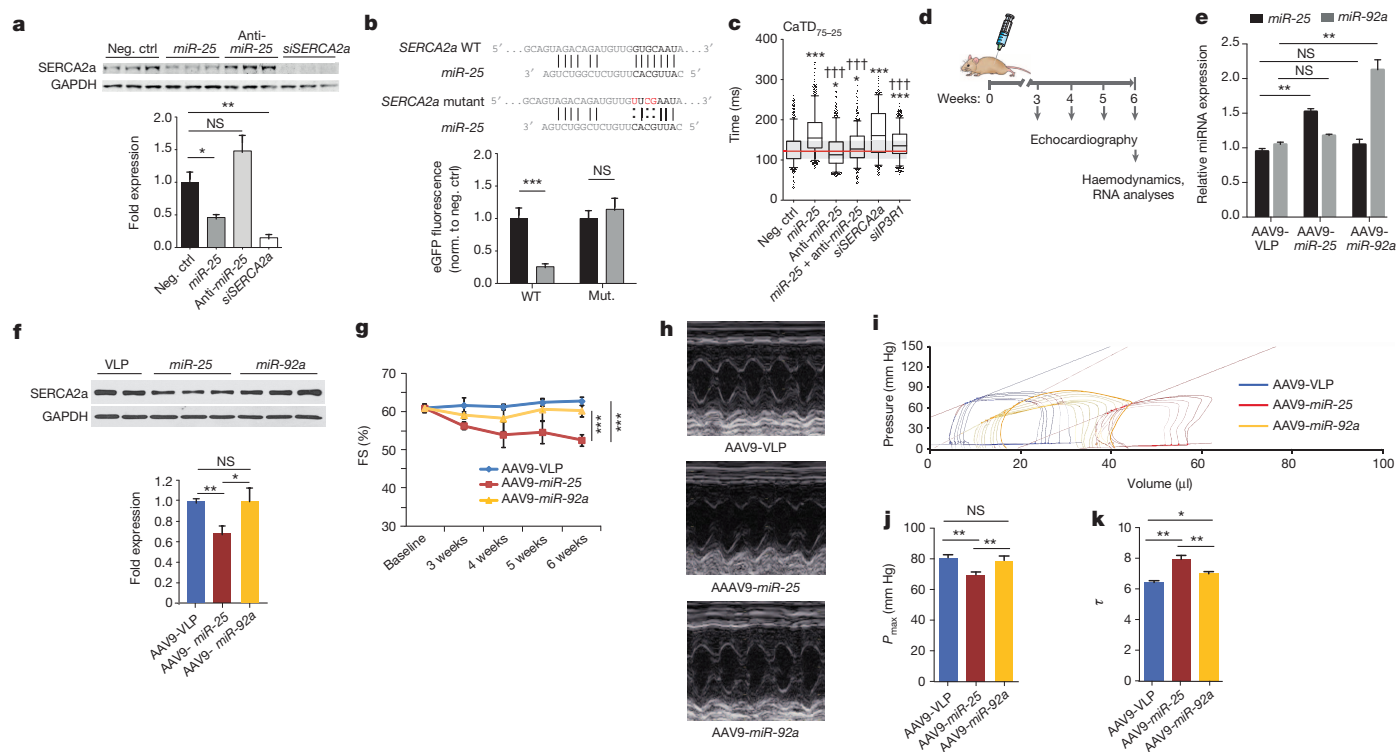
human heart failure ( $n > 550$  cells per group). Box shows 25th to 75th percentiles; whiskers show 5th and 95th percentiles; dots indicate outliers.  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ . NS, not significant (one-tailed ANOVA). **h,** Frequency distribution and log-normal curve fits for  $\text{CaTD}_{75-25}$  values from panel **f**, normalized to sample size. Both *siSerca2a* and *miR-25* increased  $\text{CaTD}_{75-25}$  values. **i,** *miR-25* is upregulated in human heart failure samples, as demonstrated by quantitative polymerase chain reaction (qPCR). Data are represented as mean  $\pm$  standard error of the mean (s.e.m.).  $**P < 0.01$  ( $n = 5$ ), Student's  $t$ -test. All replicates ( $n$ ) are biological. Primary and confirmatory screens were performed once (**d, e**); subsequent analyses were performed two (**f–h**) or three (**i**) times.

*miR-25* and mimicking the  $\sim 1.5$ – $2$ -fold decline reported for ventricular cardiomyocytes isolated from failing human hearts<sup>16,17</sup>. siRNA against IP3R1 only minimally affected  $\text{Ca}^{2+}$  transient kinetics in HL-1 cells and not at all in NRVCs ( $\text{CaTD}_{50}$ ,  $\text{CaTD}_{75-25}$  and  $V_{\text{max}}$ ; Fig. 3c and Extended Data Fig. 3), suggesting that the predominant effect of *miR-25* on  $\text{Ca}^{2+}$  transient kinetics is mediated through downregulation of SERCA2a.

We evaluated the physiological effect of blocking *miR-25*. AntagomiRs are antisense oligonucleotides modified to enhance duplex stability and have been used effectively to inhibit miRNA function *in vitro* and *in vivo*<sup>18</sup>. Anti-*miR-25* transfection alone decreased  $\text{CaTD}_{75-25}$  in HL-1 cells but had no effect in NRVCs (Fig. 3c and Extended Data Fig. 1d). Importantly, when co-transfected, anti-*miR-25* cancelled the prolonging effect of *miR-25* on  $\text{CaTD}_{75-25}$ , restoring kinetic parameters to near normal in both cell types.

We next assessed the physiological consequences of administering *miR-25* and anti-*miR-25* *in vivo*. AAV9-mediated gene transfer of *miR-25* increased the levels of *miR-25* by approximately 50% and correspondingly decreased SERCA2a levels in ventricular myocardium (Fig. 3d–f), causing a progressive decline in fractional shortening, an index of cardiac function (Fig. 3g, h). Haemodynamic analyses at the termination of the studies (6 weeks after AAV transfer) confirmed that left ventricular (LV) function had declined in *miR-25*-injected mice (Fig. 3i–k and Extended Data Fig. 4b, c). By contrast, AAV-mediated transfer of *miR-92a* (Fig. 3d, e), which has the same seed sequence (AUUGCAC) as *miR-25*, did not affect SERCA2a levels (Fig. 3f) or significantly affect LV function (Fig. 3g–k and Extended Data Fig. 4b, c), although it down-regulated integrin  $\alpha 5$ , a target involved in angiogenesis<sup>19</sup> (Extended Data Fig. 4a), despite the identical copy number of integrated AAV genomes (Extended Data Fig. 5). Together, these results point to a selective interaction between *miR-25* and *Serca2a* mRNA that is consistent with the importance of non-seed sequences for target specificity<sup>20</sup>, and also indicate that elevated *miR-25* can depress cardiac function. To





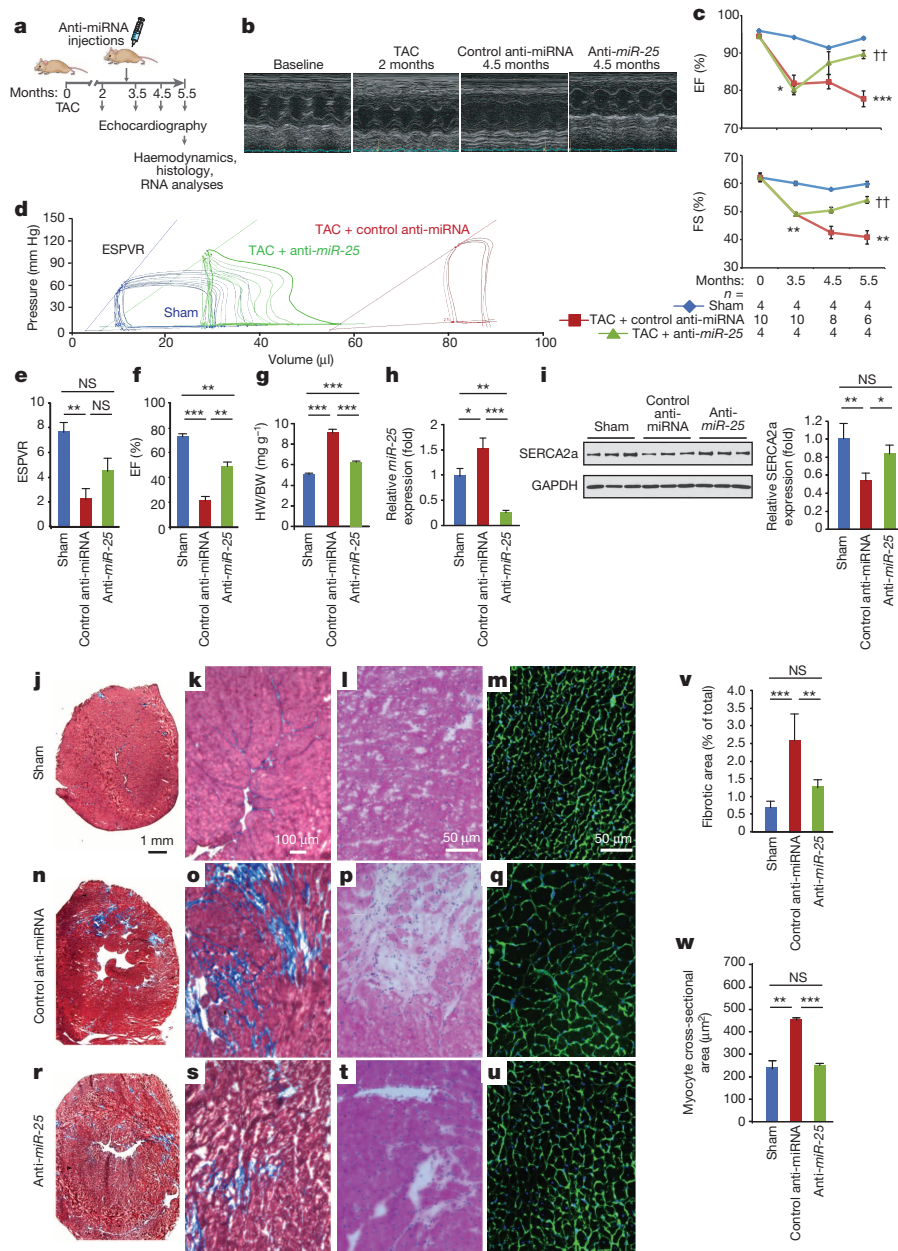
**Figure 3 | *miR-25* directly targets *SERCA2a* and regulates contractile  $\text{Ca}^{2+}$  kinetics.** **a**, Effects of *miR-25*, anti-*miR-25* and controls (scrambled sequence and *Serca2a* siRNA) on *SERCA2a* protein levels. **b**, Mutagenesis of the putative *miR-25* recognition element in the *SERCA2a* mRNA 3' UTR abolished inhibition by *miR-25*. \* $P < 0.05$ , \*\* $P < 0.01$  ( $n = 10$ ; Student's *t*-test), compared with control. Mut., mutant; Norm., normalized; WT, wild type. **c**, Anti-*miR-25* diminished the effect of *miR-25* on  $\text{CaT}_{75-25}$  in transfected HL-1 cells. siRNA against IP3R1 had a minimal effect. Extended Data Fig. 3 shows additional parameters and similar results using NRVCs. Box plots as in Fig. 1g. \*, † indicate significant differences (\* or † $P < 0.05$ , \*\* or †† $P < 0.01$ , \*\*\* or ††† $P < 0.001$ , one-tailed ANOVA) from negative control (\*) or *miR-25* (†) ( $n > 550$  cells per group). The data in panels **a–c** represent one of three

independent experiments. **d**, Protocol for AAV9-mediated cardiac gene transfer. **e**, AAV-*miR-25* ( $n = 4$ ) and AAV-*miR-92a* ( $n = 5$ ) increased levels of their respective miRNAs relative to AAV-VLP control ( $n = 3$ ). **f**, **g**, *miR-25*, but not *miR-92a*, diminished *SERCA2a* protein levels (**f**) and fractional shortening (FS; percentage) after injection ( $n = 5$  for each cohort; **g**). **h**, LV M-mode images. **i–k**, Pronounced effect of *miR-25* on pressure–volume relationship (**i**) and  $P_{\text{max}}$  (**j**) showing decreased function relative to *miR-92a* and control AAV9; and the effect on  $\tau$ , the time constant for LV relaxation, suggestive of diastolic dysfunction ( $n = 5$ , AAV9-VLP;  $n = 4$ , AAV9-*miR-25*;  $n = 5$ , AAV9-*miR-92a*). Data are represented as mean  $\pm$  s.e.m. in all panels except **c**. NS, not significant. All replicates ( $n$ ) are biological.

evaluate the selectivity and efficacy of anti-*miR-25*, we intravenously injected anti-miRNAs formulated with *in vivo*-jetPEI reagent mixture. Anti-*miR-25* injection decreased endogenous *miR-25* levels in LV myocardium of both wild-type and *SERCA2a* knockout mice relative to treatment with a control (scrambled sequence) anti-miR (Extended Data Fig. 6a), but did not affect *miR-92a* (Extended Data Fig. 7). Accordingly, anti-*miR-25* increased *SERCA2a* levels as expected in wild-type mice (Extended Data Fig. 6b). Notably, however, anti-*miR-25* did not improve the cardiac morphometric or haemodynamic parameters of *SERCA2a* knockout mice (Extended Data Fig. 8), suggesting that the  $\text{Ca}^{2+}$  pump is a critical target for *miR-25*.

Because *miR-25* decreased *SERCA2a* levels *in vitro* and *in vivo*, we evaluated whether administration of anti-*miR-25* would affect cardiac function during chronic heart failure. Mice were subjected to 3 months of TAC to chronically increase LV load and cause LV dilation before administering anti-*miR-25*. Once heart failure was established, anti-*miR-25* or a control anti-miRNA were injected intravenously and the mice were subsequently monitored for their effects on heart function (Fig. 4a). Echocardiography revealed substantial improvements in cardiac function after injection of anti-*miR-25* at 4.5 and 5.5 months after TAC, despite constant pressure overload, compared with severe deterioration in animals injected with the control anti-miRNA (Fig. 4b, c and Supplementary Table 2). Furthermore, haemodynamic analyses at the termination of the studies (5.5 months after TAC) showed substantially improved LV function in the anti-*miR-25*-injected mice, with the treatment effectively restoring the load-independent parameter

end systolic pressure volume relationship (ESPVR) and the ejection fraction (EF) to normal levels (Fig. 4d–f and Supplementary Table 3). The heart weight to body weight ratio was also stabilized (Fig. 4g). Survival of the TAC-induced heart failure animals was also improved, with 7/8 anti-*miR-25*-injected and 8/8 sham-operated animals surviving versus 7/22 of the control anti-miRNA-injected animals (Kaplan–Meier analysis, Extended Data Fig. 9), reflected by an increased probability of survival of TAC animals injected with anti-*miR-25* versus control anti-miRNA ( $P = 0.0131$ , log-rank test). Consistent with improved cardiac function, anti-*miR-25*-injected mice had significantly reduced levels of endogenous *miR-25* compared with control animals (both TAC (Fig. 4h) and sham-operated (Extended Data Fig. 6a)). Neither anti-*miR-25* nor the control anti-miRNA affected the cardiac physiology of sham-operated animals (Extended Data Fig. 8). Total and SUMOylated *SERCA2a* levels were significantly increased (Fig. 4i and Extended Data Fig. 10), indicating that anti-*miR-25* injection restored the loss of *SERCA2a* protein as well as post-translational modifications of *SERCA2a* such as SUMOylation (which improves transporter stability and ATPase activity<sup>21</sup>) despite the prior onset of heart failure. Furthermore, injection of anti-*miR-25* reduced fibrosis (Fig. 4j–l, n–p, r–t, v) and also normalized cardiomyocyte cell size (Fig. 4m, q, u, w). Given that anti-*miR-25* had no salutary effect on cardiac morphometric or haemodynamic parameters in *SERCA2a* knockout mice (Extended Data Fig. 8), did not affect expression of other miRNAs with homologous seed sequences (Extended Data Fig. 7), and that AAV9-mediated cardiac transfer of *miR-25* selectively affected *SERCA2a* (Fig. 3f), we propose that the



**Figure 4 | Inhibition of *miR-25* normalizes TAC-induced cardiac dysfunction.** **a**, Protocol for anti-*miR-25* therapy in the mouse TAC heart failure model. **b**, LV M-mode images showing dilation with control anti-*miRNA* in contrast with anti-*miR-25* injected mice. **c**, Effect on echocardiographic indices of LV function: EF and fractional shortening (FS), expressed as percentages. The number of animals initiated is  $n = 4$  (sham operated), 10 (TAC plus control anti-*miRNA*) and 4 (TAC plus anti-*miR-25*); the numbers analysed per time point are indicated and reflect deaths in the TAC plus control anti-*miRNA* group. **d–f**, Haemodynamic effects. **d**, Pressure–volume plots of treatment cohorts as indicated. Note anti-*miR-25* trend towards normalization of haemodynamic indices of end ESPVR (slope of lines in **d**, **e**), and EF (**f**).  $n = 4$  (sham operated), 4 (TAC plus control anti-*miRNA*) and 3 (TAC plus anti-*miR-25*). **g**, Heart weight (HW) to body

weight (BW) ratio.  $n = 5$  (sham), 8 (TAC plus control anti-*miRNA*) and 4 (TAC plus anti-*miR-25*). **h**, **i**, Effect of treatment on endogenous *miR-25* levels (**h**; qPCR,  $n = 4$ ) and SERCA2a protein (**i**; immunoblotting,  $n = 3$ ). **j–u**, Masson's trichrome (**j**, **k**, **n**, **o**, **r**, **s**), haematoxylin/eosin (**l**, **p**, **t**) and wheat germ agglutinin (WGA) (**m**, **q**, **u**) stained sections of hearts and LV wall. **v**, Quantified fibrotic area ( $n = 3$  (sham operated), 4 (TAC plus control anti-*miRNA*) and 5 (TAC plus anti-*miR-25*)). **w**, Average cardiomyocyte cross-sectional area ( $n = 3$  for all cohorts). For all panels, mean  $\pm$  s.e.m. \*, † indicate significant differences (\* or †  $P < 0.05$ , \*\* or ††  $P < 0.01$ , \*\*\* or †††  $P < 0.001$ , Student's *t*-test between sham and anti-*miRNA* (\*) or control anti-*miRNA* and anti-*miR-25* (†)). NS, not significant. All replicates ( $n$ ) are biological and represent one of two independent experiments.

beneficial effect of anti-*miR-25* is due to the inhibition of pathologically upregulated endogenous *miR-25* and the subsequent restoration of SERCA2a activity. To our knowledge, these data provide the first evidence that cardiac delivery of an anti-*miRNA* can directly control SERCA2a protein levels to achieve a long-term improvement in cardiac function.

These studies identified *miR-25* as a critical repressor of SERCA2a and cardiac function during heart failure. Of note, several mechanisms

beyond SERCA2a deficiency, such as changes in  $K^+$  channel density, NCX expression and myofilament sensitivity to  $Ca^{2+}$ , also contribute to slowing of the  $Ca^{2+}$  transient in advanced heart failure<sup>22</sup>. *miR-25* might also control IP3R1, and thus could be involved under conditions of inositol-1,4,5-trisphosphate sensitization, such as in response to endothelin 1, angiotensin and phenylephrine<sup>23</sup> or in local  $Ca^{2+}$  control<sup>24</sup>. NADPH oxidase 4 (NOX4) was recently shown to be downregulated

by *miR-25* in a mouse model of diabetic nephropathy<sup>25</sup>; however, NOX4 levels did not change significantly in our long-term TAC heart failure model (data not shown). Whether NOX4, which is responsible for the production of superoxide, might contribute to the salutary effects of anti-*miR-25* therapy is unclear as it is reported to confer both protective and detrimental effects<sup>26</sup>. Finally, *miR-25* was recently reported to decrease acutely after aortic constriction in mice<sup>27</sup>, in contrast with the chronic elevation reported here and in human heart failure samples, perhaps reflecting the very different stages of pathogenesis. In conclusion, our data suggest that inhibition of *miR-25* may be a novel therapeutic strategy for the treatment of heart failure.

## METHODS SUMMARY

For primary screening, HEK293 cells were co-transfected with the Ambion Pre-miR miRNA Precursor Human v.2.0 microRNA library (25 nM) and 300 ng *SERCA2a* 3' UTR target sensor plasmid (Fig. 1a) per well in 384-well plates, in triplicate.

For secondary screening,  $Ca^{2+}$  transient recordings (10 s, 33 frames per second) were acquired from HL-1 cardiomyocytes and NRVCs transfected with candidate miRNAs and loaded with Hoechst 33342 and Fluo-4 72 h after transfection, and analysed to derive cytoplasmic calcium transient kinetic parameters using a KIC instrument and software (Vala Sciences). Unresponsive or low responding cells were removed by gating. Statistical significance was determined using one-way ANOVA.

All mice were housed and treated in accordance with guidelines from the National Institutes of Health and institutional animal care and use committees, and protocols were approved by the Mount Sinai School of Medicine or Sanford-Burnham Medical Research Institute animal care and use committees.

LV samples were from explanted hearts of patients with severe heart failure, obtained at the time of cardiac transplantation. Non-failing hearts (controls) were from patients who died of cerebrovascular accidents with no evidence of contractile dysfunction by echocardiography. The 5 non-failing hearts (3 males and 2 females) had a median age of 43. The 5 heart-failure patients (4 males and 1 female) had a median age of 54 and their mean EF before cardiac transplantation was  $22 \pm 3\%$ .

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 January 2013; accepted 23 January 2014.**

**Published online 12 March 2014.**

- Bonow, R. O., Mann, D. L., Zipes, D. P. & Libby, P. *Braunwald's Heart Disease* (Saunders, 2011).
- Meyer, M. et al. Alterations of sarcoplasmic reticulum proteins in failing human dilated cardiomyopathy. *Circulation* **92**, 778–784 (1995).
- Jessup, M. et al. Calcium upregulation by percutaneous administration of gene therapy in cardiac disease (CUPID): a phase 2 trial of intracoronary gene therapy of sarcoplasmic reticulum  $Ca^{2+}$ -ATPase in patients with advanced heart failure. *Circulation* **124**, 304–313 (2011).
- Ikeda, S. et al. Altered microRNA expression in human heart disease. *Physiol. Genomics* **31**, 367–373 (2007).
- Leptidis, S. et al. A deep sequencing approach to uncover the miRNOME in the human heart. *PLoS ONE* **8**, e57800 (2013).
- Shah, A. M. & Mann, D. L. In search of new therapeutic targets and strategies for heart failure: recent advances in basic science. *Lancet* **378**, 704–712 (2011).
- Mudd, J. O. & Kass, D. A. Tackling heart failure in the twenty-first century. *Nature* **451**, 919–928 (2008).
- Greenstein, J. L. & Winslow, R. L. Integrative systems models of cardiac excitation-contraction coupling. *Circ. Res.* **108**, 70–84 (2011).
- Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
- Filipowicz, W., Bhattacharyya, S. N. & Sonenberg, N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Rev. Genet.* **9**, 102–114 (2008).
- Gurha, P. et al. Targeted deletion of microRNA-22 promotes stress-induced cardiac dilation and contractile dysfunction. *Circulation* **125**, 2751–2761 (2012).
- Kawase, Y. et al. Reversal of cardiac dysfunction after long-term expression of *SERCA2a* by gene transfer in a pre-clinical model of heart failure. *J. Am. Coll. Cardiol.* **51**, 1112–1119 (2008).
- Shirdel, E. A., Xie, W., Mak, T. W. & Jurisica, I. NAViGaTing the microneome—using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS ONE* **6**, e17429 (2011).
- Lemons, D., Maurya, M. R., Subramaniam, S. & Mercola, M. Developing microRNA screening as a functional genomics tool for disease research. *Front. Physiol.* **4**, 223 (2013).
- Cerignoli, F. et al. High throughput measurement of  $Ca^{2+}$  dynamics for drug risk assessment in human stem cell-derived cardiomyocytes by kinetic image cytometry. *J. Pharmacol. Toxicol. Methods* **66**, 246–256 (2012).
- Beuckelmann, D. J., Nabauer, M. & Erdmann, E. Intracellular calcium handling in isolated ventricular myocytes from patients with terminal heart failure. *Circulation* **85**, 1046–1055 (1992).
- Piacentino, V. III et al. Cellular basis of abnormal calcium transients of failing human ventricular myocytes. *Circ. Res.* **92**, 651–658 (2003).
- Krützfeldt, J. et al. Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
- Bonauer, A. et al. MicroRNA-92a controls angiogenesis and functional recovery of ischemic tissues in mice. *Science* **324**, 1710–1713 (2009).
- Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
- Kho, C. et al. SUMO1-dependent modulation of *SERCA2a* in heart failure. *Nature* **477**, 601–605 (2011).
- Kho, C., Lee, A. & Hajjar, R. J. Altered sarcoplasmic reticulum calcium cycling—targets for heart failure therapy. *Nature Rev. Cardiol.* **9**, 717–733 (2012).
- Higazi, D. R. et al. Endothelin-1-stimulated InsP3-induced  $Ca^{2+}$  release is a nexus for hypertrophic signaling in cardiac myocytes. *Mol. Cell* **33**, 472–482 (2009).
- Hulot, J. S. et al. Critical role for stromal interaction molecule 1 in cardiac hypertrophy. *Circulation* **124**, 796–805 (2011).
- Fu, Y. et al. Regulation of NADPH oxidase activity is associated with miRNA-25-mediated NOX4 expression in experimental diabetic nephropathy. *Am. J. Nephrol.* **32**, 581–589 (2010).
- Schmidt, H. H., Wingler, K., Kleinschnitz, C. & Disting, G. NOX4 is a Janus-faced reactive oxygen species generating NADPH oxidase. *Circ. Res.* **111**, e15–16 (2012).
- Dirkx, E. et al. Nfat and miR-25 cooperate to reactivate the transcription factor Hand2 in heart failure. *Nature Cell Biol.* **15**, 1282–1293 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank P. Aza-Blanc and F. Cerignoli; O. Kim, L. Liang and E. Kohlbrenner for their technical support; G. Christensen for providing the *SERCA2a* knockout mice; and H. el Azzouzi for TAC operations and histological sections. This work was supported by California Institute for Regenerative Medicine (RC1-000132), the National Institutes of Health (NIH) (R01HL113601, P01HL098053 and R01HL108176) and the Fondation Leducq to M.M.; by the NIH (NIH R01HL093183, R01HL088434, P20HL100396 and a Program of Excellence in Nanotechnology Contract no. HHSN26820100045C and P50HL112324) to R.J.H.; P30CA030199 and P30AR061303 for Sanford-Burnham Medical Research Institute functional genomics and cytometry. W.J.P. was supported by the Global Research Laboratory Program of the South Korean Government (M6-0605-00-0001). J.P.G.S. and P.A.F.D. were supported by the Netherlands Heart foundation and Project P1.05 LUST of the BioMedical Materials institute co-funded by the Dutch Ministry of Economic Affairs, Agriculture and Innovation. C.W. was supported by a fellowship from the Spanish National Research Council. A.v.M. was a Netherlands Heart Institute ICIN fellow.

**Author Contributions** C.W., A.R.-M., D.J., R.J.H. and M.M. conceived and designed the project following an initial concept from M.M.; C.W., D.J., A.R.-M., C.K., A.L., S.M. and A.v.M. performed experiments and analysed the data; and interpreted results with M.M. and R.J.H. W.J.P. developed reagents for post-translational modification assays on *SERCA2a*. M.M., R.J.H., C.W., A.R.-M., D.J., A.v.M., P.A.F.D. and J.P.G.S. wrote and edited the manuscript. Major funding was obtained by P.A.F.D., R.J.H. and M.M.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. ([mmmercola@sanfordburnham.org](mailto:mmmercola@sanfordburnham.org)).



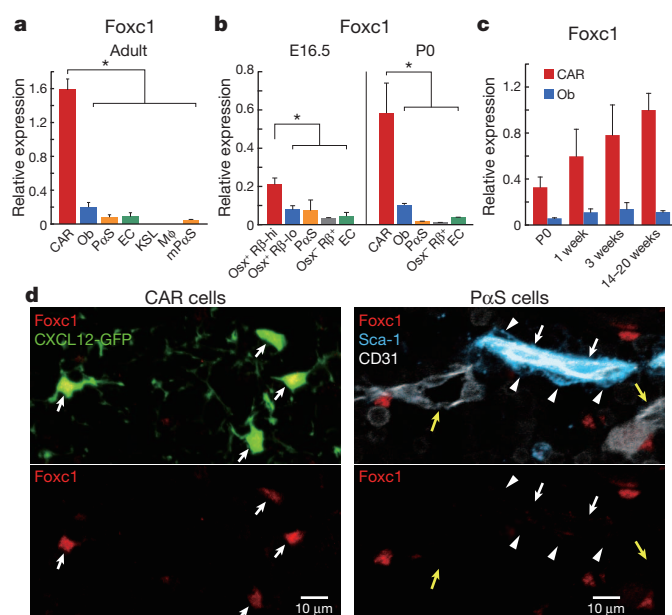
# Foxc1 is a critical regulator of haematopoietic stem/progenitor cell niche formation

Yoshiki Omatsu<sup>1,2</sup>, Masanari Seike<sup>1,2</sup>, Tatsuki Sugiyama<sup>1,2</sup>, Tsutomu Kume<sup>3</sup> & Takashi Nagasawa<sup>1,2</sup>

Haematopoietic stem and progenitor cells are maintained by special microenvironments known as niches in bone marrow<sup>1–6</sup>. Many studies have identified diverse candidate cells that constitute niches for haematopoietic stem cells in the marrow, including osteoblasts<sup>7–10</sup>, endothelial cells<sup>11–14</sup>, Schwann cells<sup>15</sup>,  $\alpha$ -smooth muscle actin-expressing macrophages<sup>16</sup> and mesenchymal progenitors such as CXCL12-abundant reticular (CAR) cells<sup>17,18</sup>, stem cell factor-expressing cells<sup>13</sup>, nestin-expressing cells<sup>19</sup> and platelet-derived growth factor receptor- $\alpha$  (PDGFR- $\alpha$ )<sup>+</sup>Sca-1<sup>+</sup>CD45<sup>–</sup>Ter119<sup>–</sup> (P $\alpha$ S) cells<sup>20</sup>. However, the molecular basis of the formation of the niches remains unclear. Here we find that the transcription factor Foxc1 is preferentially expressed in the adipo-osteogenic progenitor CAR cells essential for haematopoietic stem and progenitor cell maintenance *in vivo*<sup>5,13,18</sup> in the developing and adult bone marrow. When Foxc1 was deleted in all marrow mesenchymal cells or CAR cells, from embryogenesis onwards, osteoblasts appeared normal, but haematopoietic stem and progenitor cells were markedly reduced and marrow cavities were occupied by adipocytes (yellow adipose marrow) with reduced CAR cells. Inducible deletion of Foxc1 in adult mice depleted haematopoietic stem and progenitor cells and reduced CXCL12 and stem cell factor expression in CAR cells but did not induce a change to yellow marrow. These data suggest a role for Foxc1 in inhibiting adipogenic processes in CAR progenitors. Foxc1 might also promote CAR cell development, upregulating CXCL12 and stem cell factor expression. This study identifies Foxc1 as a specific transcriptional regulator essential for development and maintenance of the mesenchymal niches for haematopoietic stem and progenitor cells.

CAR cells express a high amount of CXCL12 and stem cell factor (SCF), which are essential for haematopoietic stem and progenitor cell (HSPC) maintenance<sup>13,14,17,21</sup>, and are largely equivalent to SCF-expressing cells<sup>14</sup> and presumably to human CD146-expressing sub-endothelial cells<sup>22</sup>. *In vivo* CAR cell ablation led to impaired HSPC maintenance<sup>18</sup>. CAR cells are adipo-osteogenic progenitors, expressing adipogenic and osteogenic genes, including osterix (Ox)<sup>18</sup>. Ox has been shown to be essential for osteoblast development and generation of both bone and marrow<sup>23</sup>. However, the molecular basis of CAR cell development remains unclear. To address this, we compared messenger RNA (mRNA) expressions of a variety of transcription factors by quantitative reverse-transcription PCR (qRT-PCR) analysis of CAR cells and alkaline phosphatase (ALP)<sup>hi</sup> osteoblasts from 14-week-old mice with the green fluorescent protein (GFP) reporter gene knocked into the *Cxcl12* locus (CXCL12-GFP mice)<sup>17</sup>. We found that the levels of Foxc1 in CAR cells were significantly higher than in osteoblasts. Foxc1 protein belongs to a family of transcription factors characterized by the presence of a forkhead box (Fox) DNA-binding domain. qRT-PCR analysis of 14- to 20-week-old CXCL12-GFP mice showed that CAR cells expressed substantially higher levels of Foxc1 than other bone marrow non-haematopoietic populations, including osteoblasts, P $\alpha$ S cells (PDGFR- $\alpha$ <sup>+</sup>Sca-1<sup>+</sup>CD31<sup>–</sup>CD45<sup>–</sup>Ter119<sup>–</sup>)<sup>20,24</sup> and endothelial

cells, as well as P $\alpha$ S adipo-osteogenic progenitors in muscle tissue (Fig. 1a). P $\alpha$ S cells are enriched for mesenchymal progenitors but distinct from CAR cells, which express PDGFR- $\alpha$  but not Sca-1 (refs 18 and 20). Haematopoietic cells, including HSPCs and macrophages, did not express Foxc1 mRNA (Fig. 1a). Single-cell RT-PCR analysis showed that all individual CAR cells expressed Foxc1 in 14-week-old CXCL12-GFP mice (50 of 50; 100%). During development, Osx-GFP<sup>+</sup>PDGFR- $\beta$ <sup>hi</sup> putative CAR progenitors from embryonic day (E)16.5 transgenic mice expressing GFP reporter gene under the control of Osx regulatory elements (Osx-GFP mice), and CXCL12-GFP<sup>+</sup> cells from newborn CXCL12-GFP mice (Extended Data Fig. 1), expressed substantially higher levels of Foxc1 than other mesenchymal populations (Fig. 1b) and the Foxc1 expression in CXCL12-GFP<sup>+</sup> cells increased over time (Fig. 1c). CXCL12-GFP<sup>+</sup> cells from newborn CXCL12-GFP mice expressed PDGFR- $\beta$ , leptin receptor (Lepr), CXCL12, SCF and peroxisome proliferator-activated receptor- $\gamma$  (PPAR $\gamma$ ), which were preferentially expressed in CAR cells<sup>14,18</sup>, but distinctly lower than adult CAR cells



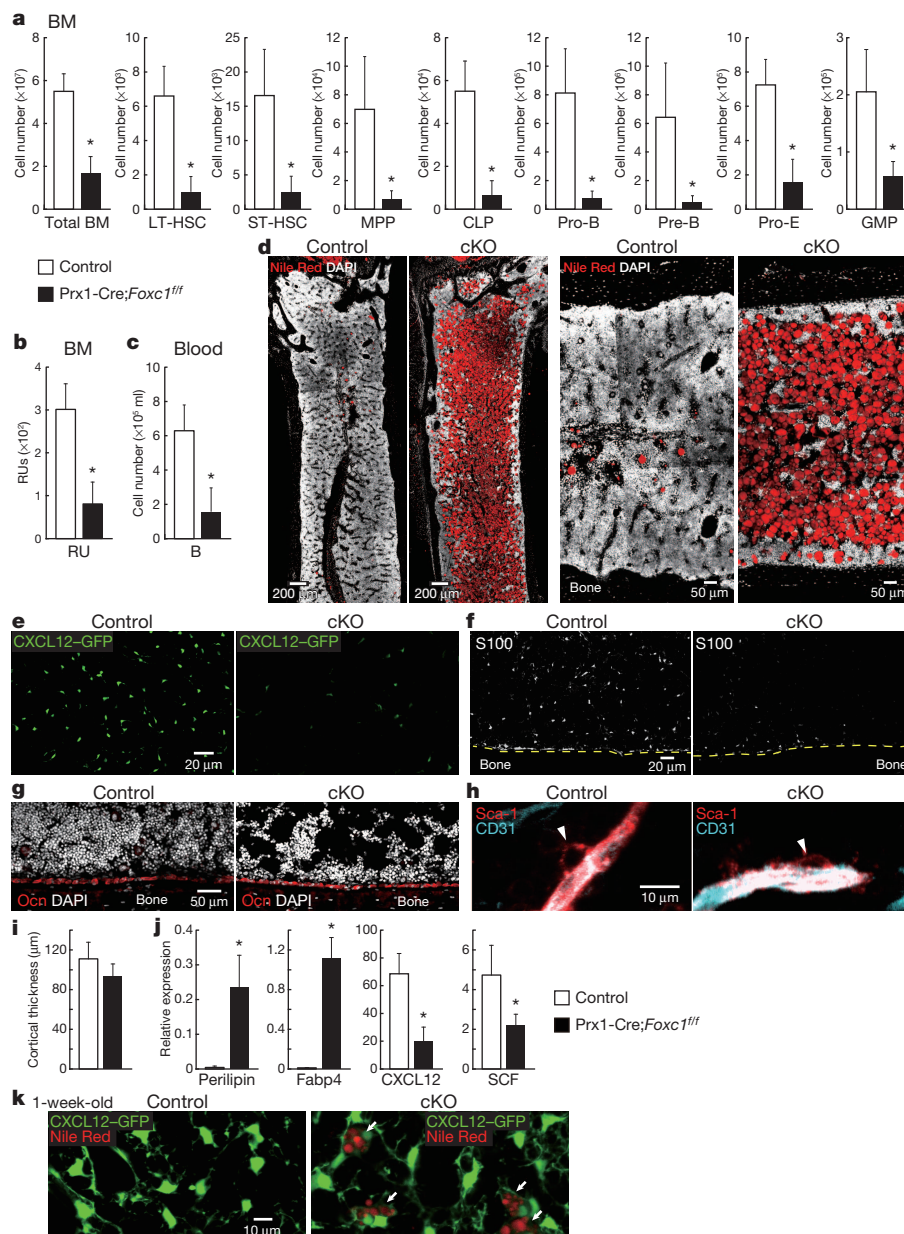
**Figure 1 | Foxc1 is preferentially expressed in CAR cells.** **a–c**, Foxc1 mRNA levels in CAR cells, osteoblasts (Ob), P $\alpha$ S cells, bone marrow endothelial cells (ECs), KSL cells, macrophages and muscle P $\alpha$ S cells from adult mice (**a**), and in Osx-GFP<sup>+</sup>PDGFR- $\beta$ <sup>hi</sup>, Osx-GFP<sup>+</sup>PDGFR- $\beta$ <sup>lo</sup>, CXCL12-GFP<sup>+</sup> cells, osteoblasts, P $\alpha$ S, Osx-GFP<sup>–</sup>PDGFR- $\beta$ <sup>+</sup>Sca-1<sup>–</sup>CD31<sup>–</sup>CD45<sup>–</sup>Ter119<sup>–</sup> (Osx<sup>–</sup>R $\beta$ <sup>+</sup>) cells and ECs from femurs of E16.5 Osx-GFP mice (**b**) or newborn (**b**, **c**), 1-week-old, 3-week-old and 14- to 20-week-old CXCL12-GFP mice (**c**) ( $n = 3$ ). **d**, Foxc1 protein was detected in CAR cells (left, arrows) but not in P $\alpha$ S cells (right, arrowheads) or ECs of arteries (right, white arrows) and sinuses (right, yellow arrows) in the marrow cavity of 2-week-old CXCL12-GFP (left) or wild-type (right) mice. Error bars, s.d. \* $P < 0.05$ .

<sup>1</sup>Department of Immunobiology and Hematology, Institute for Frontier Medical Sciences, Kyoto University, Japan. <sup>2</sup>Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST), 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan. <sup>3</sup>Feinberg Cardiovascular Research Institute, Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.

and thus were termed primordial CAR cells (Extended Data Fig. 1c, e, f). Immunohistochemical analysis of bone marrow cavities with antibodies against Foxc1 showed that Foxc1 protein was detected in CAR cells but not in P $\alpha$ S cells or endothelial cells of arteries and sinuses (Fig. 1d and Extended Data Fig. 2). These results indicate that Foxc1 is expressed preferentially in CAR cells in the developing and adult bone marrow. Foxc1 and Foxc2 are closely related Fox family members. Although Foxc2 mRNA levels were higher in CAR cells than other marrow mesenchymal populations, the differences in Foxc2 mRNA levels between CAR cells and osteoblasts were much smaller than those in Foxc1 mRNA levels (Fig. 1a–c and Extended Data Fig. 3).

Foxc1 null mutants die perinatally with haemorrhagic hydrocephalus and calvarial defects<sup>25</sup>. To determine the roles of Foxc1 in bone marrow mesenchymal niche formation, we used mice with conditional floxed Foxc1 alleles (Foxc1<sup>fl/f</sup> mice)<sup>26</sup> in conjunction with transgenic mice expressing Cre recombinase under the control of Prx1 regulatory elements (Prx1-Cre mice)<sup>27</sup>. Prx1-Cre mice express Cre recombinase in all mesenchymal cells in developing limbs and parts of the skull but not in other organs<sup>27</sup>. Prx1-Cre targeted all CAR cells, P $\alpha$ S cells and osteoblasts but not haematopoietic and endothelial cells in bone marrow<sup>20</sup>

(data not shown). Prx1-Cre;Foxc1<sup>fl/f</sup> mice were born at the expected Mendelian ratio and remained viable much longer than Foxc1 null mutants, although they showed hydrocephalus and did not survive beyond 6 weeks of age. Flow cytometric analysis showed that total haematopoietic cell counts and the numbers of the CD34<sup>+</sup>CD150<sup>+</sup>CD48<sup>−</sup> subset of c-kit<sup>+</sup>Sca-1<sup>+</sup>Lineage (Lin)<sup>−</sup> (KSL) cells, which are highly enriched for long-term repopulating haematopoietic stem cells (LT-HSCs)<sup>11</sup>, CD34<sup>+</sup>CD48<sup>−</sup> KSL short-term repopulating HSCs (ST-HSCs), Flt3<sup>+</sup> KSL multipotent progenitors (MPPs), Lin<sup>−</sup>IL-7R $\alpha$ <sup>+</sup>Flt3<sup>+</sup> common lymphoid progenitors (CLPs), c-kit<sup>+</sup>CD19<sup>+</sup>IgM<sup>−</sup> pro-B cells, c-kit<sup>+</sup>CD19<sup>+</sup>IgM<sup>−</sup> pre-B cells, c-kit<sup>+</sup>CD71<sup>+</sup>Ter119<sup>lo</sup> proerythroblasts and c-kit<sup>+</sup>Sca-1<sup>−</sup>Lin<sup>−</sup>CD34<sup>+</sup>Fc $\gamma$ R1/II<sup>hi</sup> granulocyte/macrophage progenitors were severely reduced in the bone marrow of 3-week-old Prx1-Cre;Foxc1<sup>fl/f</sup> mice compared with control animals (Fig. 2a). Additionally, the numbers of functional HSCs, including repopulating units, based on a competitive repopulation assay, and long-term *in vitro* culture-initiating cells (LTC-IC)<sup>17</sup>, based on a limiting dilution LTC-IC assay, were markedly reduced in the mutant bone marrow (Fig. 2b and Extended Data Fig. 4). Flow cytometric analysis of the bone marrow from newborn Prx1-Cre;Foxc1<sup>fl/f</sup> mice showed that the numbers of



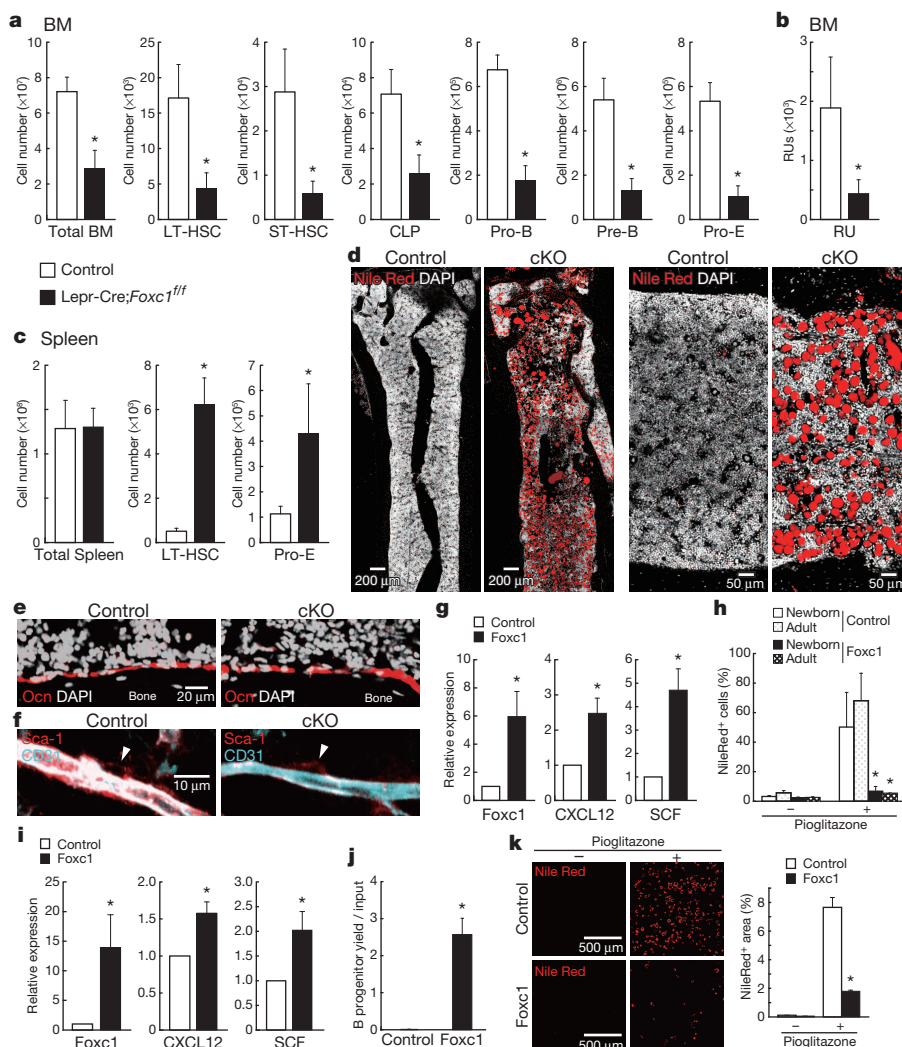
**Figure 2 | Foxc1 deletion results in yellow marrow with reduced HSPCs and CAR cells.** **a–c**, Total haematopoietic cell counts and numbers of LT-HSCs, ST-HSCs, MPPs, CLPs, pro-B cells, pre-B cells, proerythroblasts (pro-E) and granulocyte/macrophage progenitors (GMPs) in the bone marrow (BM) (**a**, **b**) and B cells in the blood (**c**) of 3-week-old control and Prx1-Cre;Foxc1<sup>fl/f</sup> mice ( $n = 5$ ). The numbers of functional HSCs were estimated using repopulating units (RUs) (**b**,  $n = 3$ ). **d–i**, Histological analysis of bone marrow from 3-week-old (**d–i**) and 1-week-old (**k**) control and Prx1-Cre;Foxc1<sup>fl/f</sup> (d, f–i) or Prx1-Cre;Foxc1<sup>fl/f</sup> CXCL12-GFP mice (**e**, **k**). **d–h**, Adipocytes (**d**, red), CAR cells (**e**, green; **f**, white), Ocn<sup>+</sup> osteoblasts (**g**, red) and Sca-1<sup>+</sup>CD31<sup>−</sup> P $\alpha$ S cells (**h**, arrowheads). **i**, The cortical thickness of femoral diaphysis ( $n = 5$ ). **j**, Perilipin, Fabp4, CXCL12 and SCF mRNA levels in residual CAR cells from 3-week-old Prx1-Cre;Foxc1<sup>fl/f</sup> and control mice ( $n = 5$ ). **k**, CAR cells (green) with lipid inclusions (red) differentiated into preadipocytes (arrows). cKO, Prx1-Cre;Foxc1<sup>fl/f</sup> mice. Error bars, s.d. \* $P < 0.05$ .

CD150<sup>+</sup>CD48<sup>-</sup> KSL cells (HSCs), pro-B cells, pre-B cells and proerythroblasts were reduced although to a lesser extent than juvenile mutants (Extended Data Fig. 5). In peripheral blood of *Prx1-Cre; Foxc1<sup>fl/fl</sup>* mice, B cells were decreased but primitive haematopoietic cells were not increased (Fig. 2c and data not shown).

Surprisingly, histological analysis of 3-week-old *Prx1-Cre; Foxc1<sup>fl/fl</sup>* mice showed that most marrow volume was occupied by adipocytes with few haematopoietic cells, and haematopoietically active marrow was restricted to the subendosteal region near the bone surface (Fig. 2d). Immunohistochemical analysis of CXCL12-GFP mice with antibodies against cytoplasmic protein S100 showed that CAR cells could be identified as S100<sup>+</sup> cells in the marrow cavity (Extended Data Fig. 6). Immunohistochemical analysis showed that the numbers of CXCL12-GFP<sup>+</sup> or S100<sup>+</sup> CAR cells were severely reduced (Fig. 2e, f) but Ocn<sup>+</sup> osteoblasts lining the endosteal bone surface and P+S cells (Fig. 2g, h and Extended Data Fig. 7a) appeared normal in the bone marrow of *Prx1-Cre*-targeted mice compared with control animals. Consistent with this, *Prx1-Cre; Foxc1<sup>fl/fl</sup>* mice had normal or slightly reduced cortical thickness of the femoral diaphysis (Fig. 2i). qRT-PCR analysis showed that residual CAR cells in *Prx1-Cre*-targeted mice had higher expression levels of adipocyte markers, including perilipin and fatty acid-binding protein 4 (Fabp4) and lower levels of CXCL12 and SCF expression than control animals (Fig. 2j). Histological analysis of bone marrow showed that primordial CAR cells appeared normal and few adipocytes were observed in newborn *Prx1-Cre; Foxc1<sup>fl/fl</sup>* CXCL12-GFP mice (data not

shown); however, CAR cells that differentiated into multivacuolar lipid-containing preadipocytes were detected in the mutants at 1 week of age (Fig. 2k). Thus, in the absence of Foxc1 during postnatal CAR cell development, most primordial CAR cells might differentiate to become adipocytes, which could not function as a HSPC niche<sup>28</sup>, and thus the numbers of CAR cells, which function as niches for HSC homing and HSPC maintenance, would be severely reduced, resulting in a reduction in HSPCs.

To analyse the effects of CAR cell-specific disruption of Foxc1, we crossed *Foxc1<sup>fl/fl</sup>* mice with *Lepr-Cre* knock-in mice. *Lepr-Cre* targeted approximately 72% and 92% of CAR cells but not other bone marrow cells in the 3-week-old and 14-week-old mutants, respectively<sup>13</sup> (data not shown). *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice were viable and did not show hydrocephalus. Flow cytometric analysis showed that total haematopoietic cell counts and the numbers of LT-HSCs, ST-HSCs, CLPs, pro-B cells, pre-B cells and proerythroblasts were severely reduced in the bone marrow of 14- to 18-week-old *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice compared with control animals (Fig. 3a). Additionally, the numbers of functional LT-HSCs, including repopulating units and LTC-ICs, were markedly reduced in the mutant marrow (Fig. 3b and Extended Data Fig. 8). In contrast, LT-HSCs and proerythroblasts were increased in the spleen but not in the peripheral blood of the mutants (Fig. 3c and data not shown), suggesting that reduction in HSPCs in the marrow might result in extramedullary haematopoiesis. Histological analysis of 18-week-old *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice showed that most marrow volume was occupied by



**Figure 3 | Essential roles of Foxc1 in CAR cells.** **a–c**, Total haematopoietic cell counts and the numbers of LT-HSCs, ST-HSCs, CLPs, pro-B cells, pre-B cells and pro-E cells (**a**,  $n = 4$ ) and functional HSCs estimated using repopulating units (**b**,  $n = 3$ ) in bone marrow, and LT-HSCs and pro-E cells in spleen (**c**) from control and *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice ( $n = 4$ ). **d–f**, Adipocytes (**d**, red), Ocn<sup>+</sup> osteoblasts (**e**, red) and Sca-1<sup>+</sup>CD31<sup>-</sup> P+S cells (**f**, arrowheads) in the marrow from control and *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice. cKO, *Lepr-Cre; Foxc1<sup>fl/fl</sup>* mice. **g–k**, Analysis of primordial (**g**, **h**) and adult (**h**) CAR cells or OP9 cells (**i–k**) transduced with control or Foxc1-expression construct. Foxc1, CXCL12 and SCF mRNA levels (**g**, **i**). Percentages of Nile-Red-positive cells (**h**) or Nile-Red-positive area (**k**, right) in pioglitazone-treated cultures ( $n = 3$ ). **j**, The numbers of B220<sup>+</sup>CD11b<sup>-</sup> IgM<sup>-</sup> B-cell progenitors generated from Lin<sup>-</sup> primitive haematopoietic progenitors in culture on OP9 cells transduced with control or Foxc1-expression construct ( $n = 3$ ). Error bars, s.d. \* $P < 0.05$ .



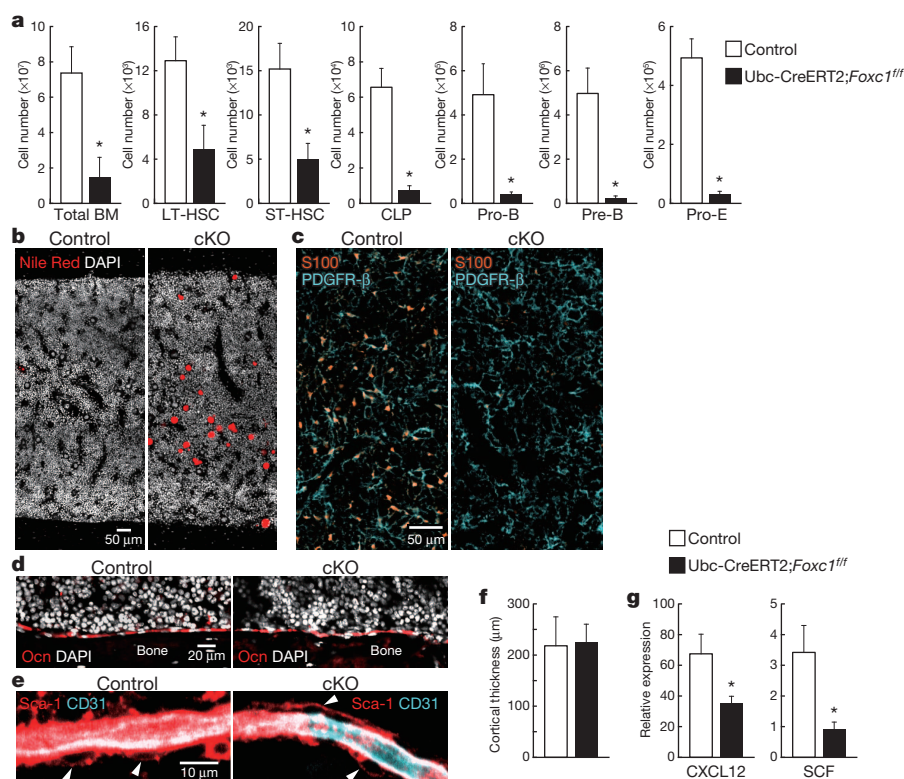
adipocytes (Fig. 3d) but endosteal osteoblasts, P $\alpha$ S cells and cortical thickness of the femoral diaphysis appeared normal (Fig. 3e, f, Extended Data Fig. 7b and data not shown). These results support the idea that Foxc1 of CAR cells has an essential role in inhibition of their differentiation to adipocytes and HSPC niche formation. In contrast, HSPC maintenance and haematopoiesis were unimpaired in 14- to 18-week-old *Foxc1*<sup>fl/fl</sup> mice in conjunction with Tie2-Cre transgenic mice, in which the *Foxc1* gene was deleted in endothelial and haematopoietic cells (Extended Data Fig. 9), indicating that Foxc1 in endothelial and haematopoietic cells is not required for haematopoiesis.

To confirm the *in vivo* function of Foxc1 in CAR cell development, we examined the effects of Foxc1 expression in CAR cells and an adipocyte progenitor cell line, OP9, which was derived from newborn op/op mouse calvaria<sup>29</sup>. Enforced expression of Foxc1 by retroviral infection in sorted primordial and adult CAR cells increased CXCL12 and SCF mRNA expression as assessed by qRT-PCR (Fig. 3g and data not shown) and markedly decreased adipogenic potential in the culture (Fig. 3h). OP9 cells expressed Foxc1 at very low levels and had the potential to differentiate into adipocytes in an *in vitro* culture (Fig. 3i, k). Enforced expression of Foxc1 in OP9 cells increased CXCL12 and SCF mRNA expression and ability to support B lymphopoiesis, and led to a loss of adipogenic potential in the culture (Fig. 3i–k).

To determine the role of Foxc1 in the maintenance of HSPC niches, we induced deletion of the *Foxc1* gene in adults by using *Foxc1*<sup>fl/fl</sup> mice in conjunction with Ubc-CreERT2 transgenic mice, allowing recombination of the floxed *Foxc1* allele upon tamoxifen treatment (Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice). Total haematopoietic cell counts and the numbers of LT-HSCs, ST-HSCs, MPPs, CLPs, pro-B cells, pre-B cells and proerythroblasts were severely reduced in the bone marrow of 16- to 22-week-old tamoxifen-treated Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice compared with control animals (Fig. 4a and data not shown). In contrast, LT-HSCs were increased in the spleen of the mutants (data not shown), suggesting that Foxc1 plays a bone-marrow-specific role in maintaining HSCs in adults. Histological analysis of bone marrow showed that only small numbers of adipocytes were observed (Fig. 4b) and the numbers of S100<sup>+</sup>PDGFR- $\beta$ <sup>+</sup> cells were reduced (Fig. 4c) but processes of PDGFR- $\beta$ <sup>+</sup> cells forming

a network appeared normal in tamoxifen-treated Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice compared with control animals (Fig. 4c). These results indicate that many CAR cells, whose niche functions were impaired, were present in the mutants. Immunohistochemical analysis showed that endothelial cells, osteoblasts, P $\alpha$ S cells (Fig. 4d, e and data not shown) and cortical thickness of the femoral diaphysis (Fig. 4f) appeared normal in the mutant bone marrow. qRT-PCR analysis showed that residual CAR cells had lower levels of CXCL12 and SCF expression, but slightly higher levels of perilipin and Fabp4 expression in tamoxifen-treated Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice compared with control animals (Fig. 4g and data not shown). These results indicate that Foxc1 is essential for maintenance of CXCL12 and SCF expression in CAR cells and HSPC niche functions of adult marrow microenvironments. Although enforced expression of Foxc1 in adult CAR cells inhibited adipogenesis *in vitro*, bone marrow adipogenesis remained blocked after induction of Foxc1 deletion in adults. Further studies will be needed to delineate a cell-intrinsic or non-cell-intrinsic mechanism by which adipogenesis is blocked in adult marrow.

We have shown that Foxc1 is essential for CAR cell development. During postnatal development, primordial CAR cells begin to express adipogenic genes (Extended Data Fig. 1f) and upregulate the expression of Foxc1, which might promote CAR cell development upregulating CXCL12 and SCF expression and subsequently inhibit the differentiation of developing CAR cells into adipocytes. Our findings, together with the knowledge that Foxn1 is essential for the development of thymic epithelium, which supports T-cell production<sup>30</sup>, establish that Fox proteins play critical but distinct roles in the formation of different haematopoietic microenvironments. Clinically, it may be possible that Foxc1 or its activators can be used therapeutically to drive the formation of haematopoietic niches from non-niche cells, such as primary dermal fibroblasts or induced pluripotent stem (iPS)-derived mesenchymal cells *in vitro*, or from impaired marrow microenvironments *in vivo*. This study provides a valuable starting point in understanding the molecular mechanisms of HSPC niche formation and new insights into features of haematopoietic microenvironments and mesenchymal progenitors.



**Figure 4 | Foxc1 is essential for maintenance of HSPC niche functions in adult bone marrow.** Bone marrow from 16- to 22-week-old (**a**, **f**, **g**) or 18-week-old (**b–e**) tamoxifen-treated control and Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice was analysed. **a**, Total haematopoietic cell counts and the numbers of LT-HSCs, ST-HSCs, CLPs, pro-B cells, pre-B cells and pro-E cells ( $n = 4$ ). **b–e**, Adipocytes (**b**, red), S100<sup>+</sup> CAR cells (**c**, orange), processes of PDGFR- $\beta$ <sup>+</sup> cells (**c**, blue), Ocn<sup>+</sup> osteoblasts (**d**, red) and Sca-1<sup>+</sup>CD31<sup>-</sup> P $\alpha$ S cells (**e**, arrowheads). **f**, The cortical thickness of femoral diaphysis ( $n = 5$ ). **g**, CXCL12 and SCF mRNA levels in CAR cells ( $n = 4$ ). cKO, Ubc-CreERT2; *Foxc1*<sup>fl/fl</sup> mice. Error bars, s.d. \* $P < 0.05$ .

## METHODS SUMMARY

*Foxc1<sup>fl/f</sup>* mice<sup>26</sup> and CXCL12–GFP knock-in mice<sup>17</sup> have been described previously. Details of procedures and reagents are described in Methods.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 1 July 2013; accepted 22 January 2014.**

**Published online 2 March 2014.**

- Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598–611 (2008).
- Li, L. & Clevers, H. Coexistence of quiescent and active adult stem cells in mammals. *Science* **327**, 542–545 (2010).
- Ehninger, A. & Trumpp, A. The bone marrow stem cell niche grows up: mesenchymal stem cells and macrophages move in. *J. Exp. Med.* **208**, 421–428 (2011).
- Bianco, P. Bone and the hematopoietic niche: a tale of two stem cells. *Blood* **117**, 5281–5288 (2011).
- Nagasawa, T., Omatsu, Y. & Sugiyama, T. Control of hematopoietic stem cells by the bone marrow stromal niche: the role of reticular cells. *Trends Immunol.* **32**, 315–320 (2011).
- Mercier, F. E., Ragu, C. & Scadden, D. T. The bone marrow at the crossroads of blood and immunity. *Nature Rev. Immunol.* **12**, 49–60 (2011).
- Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
- Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
- Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161 (2004).
- Sugimura, R. *et al.* Noncanonical Wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).
- Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
- Butler, J. M. *et al.* Endothelial cells are essential for the self-renewal and repopulation of Notch-dependent hematopoietic stem cells. *Cell Stem Cell* **6**, 251–264 (2010).
- Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature* **481**, 457–462 (2012).
- Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
- Yamazaki, S. *et al.* Nonmyelinating Schwann cells maintain hematopoietic stem cell hibernation in the bone marrow niche. *Cell* **147**, 1146–1158 (2011).
- Ludin, A. *et al.* Monocytes-macrophages that express  $\alpha$ -smooth muscle actin preserve primitive hematopoietic cells in the bone marrow. *Nature Immunol.* **13**, 1072–1082 (2012).
- Sugiyama, T., Kohara, H., Noda, M. & Nagasawa, T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* **25**, 977–988 (2006).
- Omatsu, Y. *et al.* The essential functions of adipo-osteogenic progenitors as the hematopoietic stem and progenitor cell niche. *Immunity* **33**, 387–399 (2010).
- Mendez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
- Greenbaum, A. *et al.* CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* **495**, 227–230 (2013).
- Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
- Sacchetti, B. *et al.* Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell* **131**, 324–336 (2007).
- Chan, C. K. *et al.* Endochondral ossification is required for haematopoietic stem-cell niche formation. *Nature* **457**, 490–494 (2009).
- Morikawa, S. *et al.* Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. *J. Exp. Med.* **206**, 2483–2496 (2009).
- Kume, T. *et al.* The forkhead/winged helix gene Mf1 is disrupted in the pleiotropic mouse mutation congenital hydrocephalus. *Cell* **93**, 985–996 (1998).
- Sasman, A. *et al.* Generation of conditional alleles for *Foxc1* and *Foxc2* in mice. *Genesis* **50**, 766–774 (2012).
- Logan, M. *et al.* Expression of Cre Recombinase in the developing mouse limb bud driven by a *Prx1* enhancer. *Genesis* **33**, 77–80 (2002).
- Naveiras, O. *et al.* Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. *Nature* **460**, 259–263 (2009).
- Kodama, H., Nose, M., Niida, S. & Nishikawa, S. Involvement of the c-kit receptor in the adhesion of hematopoietic stem cells to stromal cells. *Exp. Hematol.* **22**, 979–984 (1994).
- Nehls, M. *et al.* Two genetically separable steps in the differentiation of thymic epithelium. *Science* **272**, 886–889 (1996).

**Acknowledgements** We appreciate the technical assistance provided by K. Kawaguchi and G. Kondoh, and thank I. Sasagawa for secretarial assistance. This research was supported by JST, CREST and the Ministry of Education, Culture, Sports, Science and Technology (MEXT)/Japan Society for the Promotion of Science (JSPS) KAKENHI.

**Author Contributions** Y.O. and T.N. designed and performed the experiments, analysed the data and prepared the paper. T.N. supervised the study. M.S. and T.S. performed the experiments. T.K. contributed materials and tools. All authors discussed results and edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.N. ([tnagasa@frontier.kyoto-u.ac.jp](mailto:tnagasa@frontier.kyoto-u.ac.jp)).

# Cell-cycle-regulated activation of Akt kinase by phosphorylation at its carboxyl terminus

Pengda Liu<sup>1</sup>, Michael Begley<sup>2,3</sup>, Wojciech Michowski<sup>4</sup>, Hiroyuki Inuzuka<sup>1</sup>, Miriam Ginzberg<sup>5</sup>, Daming Gao<sup>1</sup>, Peiling Tsou<sup>2,3</sup>, Wenjian Gan<sup>1</sup>, Antonella Papa<sup>1,2,6</sup>, Byeong Mo Kim<sup>8</sup>, Lixin Wan<sup>1</sup>, Amrik Singh<sup>7</sup>, Bo Zhai<sup>5</sup>, Min Yuan<sup>2</sup>, Zhiwei Wang<sup>1,†</sup>, Steven P. Gygi<sup>5</sup>, Tae Ho Lee<sup>8</sup>, Kun-Ping Lu<sup>2</sup>, Alex Toker<sup>1</sup>, Pier Paolo Pandolfi<sup>1,2,6</sup>, John M. Asara<sup>2</sup>, Marc W. Kirschner<sup>3</sup>, Piotr Sicinski<sup>4</sup>, Lewis Cantley<sup>2,3,†</sup> & Wenyi Wei<sup>1</sup>

**Akt, also known as protein kinase B, plays key roles in cell proliferation, survival and metabolism. Akt hyperactivation contributes to many pathophysiological conditions, including human cancers<sup>1–3</sup>, and is closely associated with poor prognosis and chemo- or radio-therapeutic resistance<sup>4</sup>. Phosphorylation of Akt at S473 (ref. 5) and T308 (ref. 6) activates Akt. However, it remains unclear whether further mechanisms account for full Akt activation, and whether Akt hyperactivation is linked to misregulated cell cycle progression, another cancer hallmark<sup>7</sup>. Here we report that Akt activity fluctuates across the cell cycle, mirroring cyclin A expression. Mechanistically, phosphorylation of S477 and T479 at the Akt extreme carboxy terminus by cyclin-dependent kinase 2 (Cdk2)/cyclin A or mTORC2, under distinct physiological conditions, promotes Akt activation through facilitating, or functionally compensating for, S473 phosphorylation. Furthermore, deletion of the cyclin A2 allele in the mouse olfactory bulb leads to reduced S477/T479 phosphorylation and elevated cellular apoptosis. Notably, cyclin A2-deletion-induced cellular apoptosis in mouse embryonic stem cells is partly rescued by S477D/T479E-Akt1, supporting a physiological role for cyclin A2 in governing Akt activation. Together, the results of our study show Akt S477/T479 phosphorylation to be an essential layer of the Akt activation mechanism to regulate its physiological functions, thereby providing a new mechanistic link between aberrant cell cycle progression and Akt hyperactivation in cancer.**

Using single live-cell imaging<sup>8</sup>, we found that Akt activation fluctuated across the cell cycle, inversely correlating with Cdt1 abundance<sup>9</sup> (Fig. 1a and Supplementary Fig. 1a). Statistical analysis of immunostained HeLa cells further showed that Akt-pS473 has a similar periodic feature as geminin<sup>9,10</sup> (Fig. 1b). Notably, in several cancer cell lines (Fig. 1c, d and Supplementary Fig. 1b, c), Akt phosphorylation, but not total Akt abundance, fluctuated across the cell cycle. The periodic Akt phosphorylation mirrored the expression pattern of cyclin A2, the predominant mammalian cyclin A isoform<sup>11</sup>, during cell cycle progression (Fig. 1c, d). Moreover, acute depletion of cyclin A2 or Cdk2, but not cyclin E, resulted in decreased Akt phosphorylation, with no significant impact on phosphorylation of Akt upstream kinases PDK1 and mTORC2 (Fig. 1d). This prompted us to evaluate whether Cdk2/cyclin A directly regulates Akt activation in a phosphorylation-dependent manner during the cell cycle<sup>12</sup>.

In support of Akt as a Cdk2/cyclin A substrate, Akt isoforms interacted with cyclin A2 (Fig. 2a and Supplementary Fig. 2a). Furthermore, we identified four 'RXL' cyclin A-binding motifs<sup>13</sup> in Akt1 (Fig. 2b), all of which are evolutionarily conserved (Supplementary Fig. 2b). Mutation of R76CL or R273DL, and to a lesser extent, R200VL or R370TL (RXL to AXA) attenuated Akt1 interaction with cyclin A2 (Fig. 2c), and reduced

Akt1 activity (Supplementary Fig. 2c). Consistently, depleting cyclin A2 or Cdk2 (Supplementary Fig. 2d–f) led to a significant reduction in Akt phosphorylation. More importantly, either acute treatment with Cdk2 inhibitors (Fig. 2d) or deletion of the cyclin A2 allele in cyclin A2<sup>fl/fl</sup> primary mouse embryonic fibroblasts (MEFs) (Fig. 2e) led to a marked decrease in Akt phosphorylation without a significant perturbation of cell cycle progression (Fig. 2d and ref. 14), excluding a possible indirect cell cycle effect on Akt phosphorylation by inhibiting Cdk2/cyclin A.

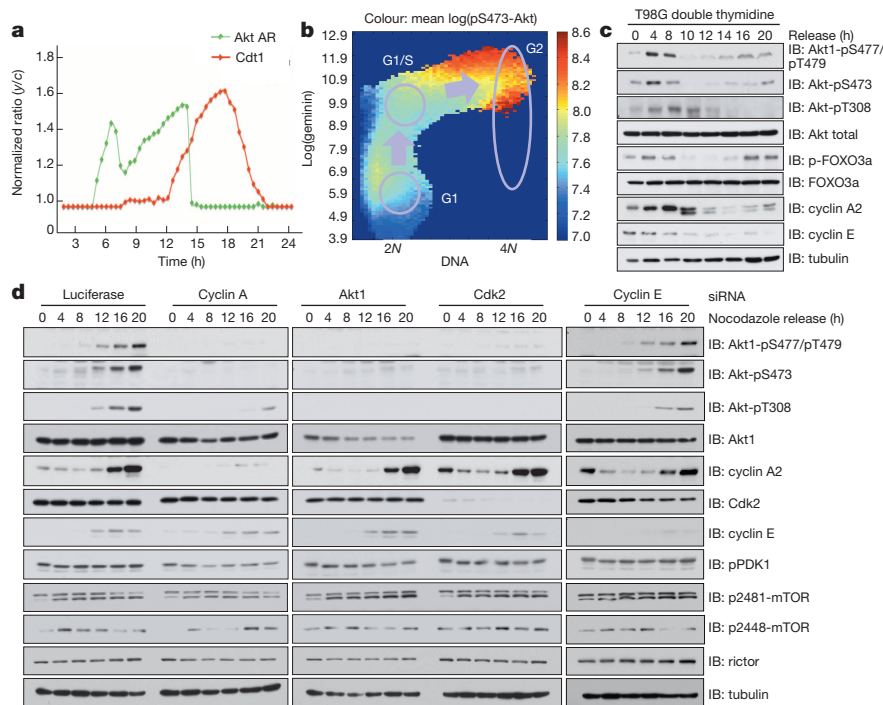
Notably, deletion of cyclin A2, but not cyclin A1 or cyclin E1/E2 alleles, caused a significant decrease of Akt phosphorylation (Fig. 2f, g), whereas conversely, ectopic expression of cyclin A2 (Fig. 2h and Supplementary Fig. 3a) resulted in elevated Akt phosphorylation coupled with enhanced *in vitro* anchorage-independent growth (Supplementary Fig. 3b, c). Moreover, depletion of Cdh1, the E3 ligase that controls cyclin A turnover<sup>15</sup>, resulted in increased cyclin A2 abundance and elevated Akt phosphorylation, leading to enhanced *in vitro* anchorage-independent growth (Supplementary Fig. 3d, e) and *in vivo* tumour formation (Fig. 2i and Supplementary Fig. 3f–h). More importantly, increased Akt phosphorylation and tumorigenicity by depleting Cdh1 could be partly reversed by extra depletion of cyclin A2 (Fig. 2j and Supplementary Fig. 3i–l). Collectively, these results support Cdk2/cyclin A2 as a major physiological kinase that governs Akt phosphorylation and oncogenic functions.

Notably, Cdk2/cyclin A directly phosphorylated Akt1 *in vitro* on its carboxy (C)-terminal region (Supplementary Fig. 4a, b). Serial truncations showed Cdk2/cyclin A phosphorylation sites in the last four evolutionarily conserved residues and subsequent mutageneses pinpointed both S477 and T479 as Cdk2/cyclin A sites (Supplementary Fig. 4c–e). Similarly, Cdk2/cyclin A phosphorylated Akt2–S478 (Supplementary Fig. 4f). Interestingly, mutation of G478 to proline (G478P) to mimic the canonical Cdk2 'SP/TP' phospho-motif<sup>16,17</sup>, or to other bulky amino acids (L/W/R), did not significantly affect Cdk2/cyclin A-mediated Akt phosphorylation (Supplementary Fig. 4g). Conversely, C-terminal addition of an  $\alpha$ -helix<sup>18</sup> (Supplementary Fig. 4g) or a green fluorescent protein (GFP) (Supplementary Fig. 4h) reduced Cdk2/cyclin A-mediated phosphorylation of the engineered non-tail version of S477G478-, but not S477P478-Akt1. These data indicate that S477/T479 may belong to a new class of Cdk2/cyclin A phospho-motifs where relative structural flexibility at the Akt1 C terminus might override the requirement of an adjacent proline for Cdk2-mediated phosphorylation of canonical TP/SP sites<sup>16,17</sup> typically buried within defined structures.

Furthermore, mass spectrometry analyses confirmed Akt1 S477 and T479 phosphorylation<sup>19</sup> (Supplementary Fig. 5a–d). To gain further mechanistic insights, we developed phospho-specific antibodies that recognize pS477/pT479-Akt1 (Supplementary Fig. 6a–f), pS477-Akt1 or pT479-Akt1 (Supplementary Fig. 7a–h) *in vivo*. Owing to the large

<sup>1</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. <sup>2</sup>Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. <sup>3</sup>Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>6</sup>Cancer Genetics Program and Division of Genetics, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA. <sup>7</sup>Cell Signaling Technology, Danvers, Massachusetts 01923, USA. <sup>8</sup>Division of Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. <sup>†</sup>Present addresses: The Cyrus Tang Hematology Center, Jiangsu Institute of Hematology, the First Affiliated Hospital, Soochow University, Suzhou 215123, China (Z.W.); Cancer Center at Weill Cornell Medical College and NewYork-Presbyterian Hospital, New York, New York 10065, USA (L.C.).





**Figure 1 | Akt activity fluctuated during the cell cycle and mirrored the periodic cyclin A expression pattern.** **a**, Single live-cell imaging of the Akt activity reporter (green, EGFP-Akt AR (activity reporter)) and the cell cycle marker Cdt1 (red, mCherry-Cdt1) transiently expressed in non-synchronized HeLa cells. **b**, A representative heat map of cell-cycle-dependent Akt1-pS473 as a function of DNA content (x axis) and geminin expression as an indicator of cell cycle stages (y axis). Blue, low Akt-pS473; red, high Akt-pS473. Number of cells was more than 500. **c**, Akt phosphorylation

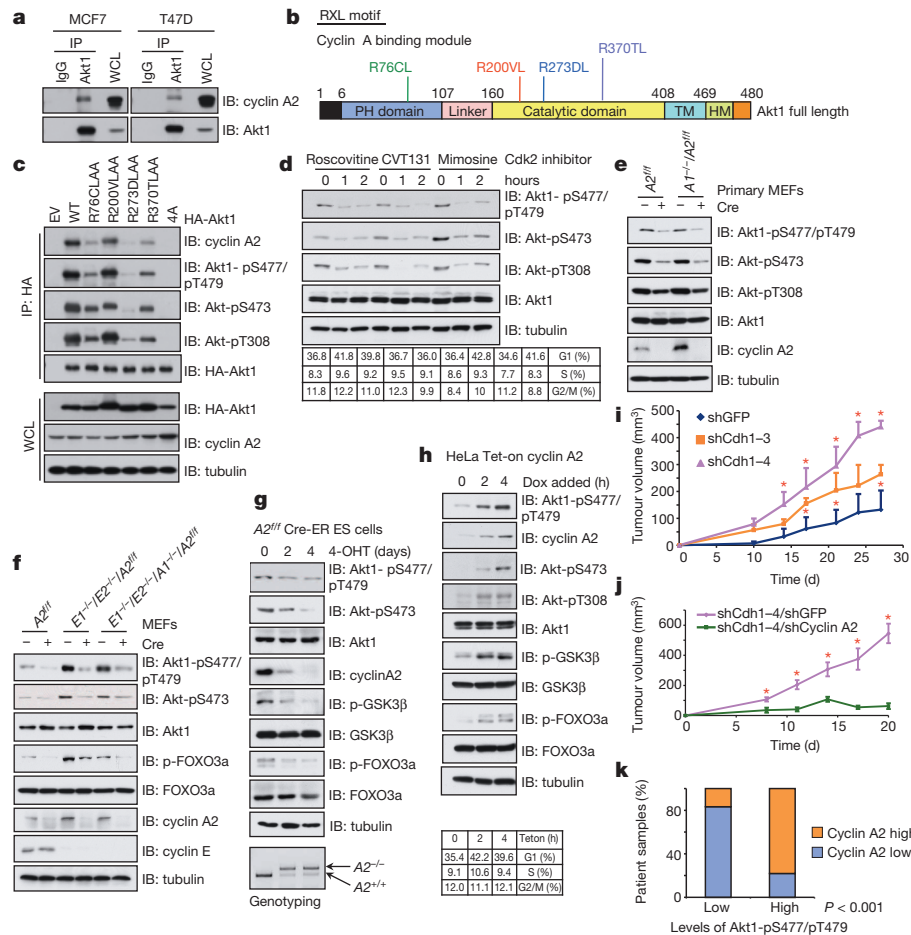
fluctuated across the cell cycle. Immunoblot (IB) of whole-cell lysates (WCLs) derived from T98G cells synchronized by double thymidine block and released to normal cell cycle for the indicated periods. **d**, Depletion of cyclin A, but not cyclin E, led to reduced Akt phosphorylation across the cell cycle. Immunoblot of whole-cell lysates derived from HeLa cells synchronized by nocodazole and released for the indicated periods. Where indicated, short interfering RNA (siRNA) oligonucleotides were transfected into cells 24 h before synchronization.

consistency among three antibodies under our experimental conditions, we focused on examining the correlation of Akt1-pS477/pT479 and Akt activation in the remainder of the studies. As with pS473-Akt1, Akt1 tail phosphorylation fluctuated during the cell cycle (Fig. 1c, d and Supplementary Fig. 1b, c), and was subjected to regulation by Cdk2/cyclin A (Fig. 2d–h). Furthermore, a positive correlation between cyclin A2 expression and Akt1-pS477/pT479 was observed in breast cancer patient samples (Fig. 2k and Supplementary Fig. 8a, b). More importantly, similar to the reported phosphorylation of Akt-S473 by several upstream kinases dependent on upstream stimuli<sup>15,20,21</sup>, Akt1-pS477/pT479 could be mediated by mTOR (Supplementary Figs 9a–c, 10a–g and 11a–c) or DNAPK (Supplementary Fig. 9d, e), in addition to Cdk2/cyclin A. In keeping with this notion, elevated Akt1-pS477/pT479 was detected in *Pten* heterozygous MEFs (Supplementary Fig. 10h). Interestingly, Akt1-pS477/pT479 was negatively regulated by PTEN in mice in a tissue-specific manner and largely with Akt-pS473 (Supplementary Fig. 10i).

Moreover, inactivation of mTORC2 by depleting Rictor led to a more dramatic reduction of Akt1-pS477/pT479 in response to insulin (Supplementary Fig. 11a) than under synchronized cell cycle conditions (Supplementary Fig. 12a). Conversely, depletion of Cdk2, but not Rictor (Supplementary Fig. 12a, b), resulted in a more robust reduction in Akt1-pS477/pT479 across the cell cycle. These findings suggest that Akt1-pS477/pT479 is possibly mediated by Cdk2/cyclin A, mTORC2 or DNAPK, under cell cycle progression, growth factor stimulation or DNA damaging conditions, respectively (Supplementary Fig. 13a, b). Notably, depletion of Cdk2, cyclin A2 or Rictor in human primary foreskin fibroblasts all resulted in reduced Akt1-pS477/pT479 (Supplementary Fig. 14a–c), highlighting both Cdk2/cyclin A and mTORC2 as upstream physiological kinases governing Akt1-pS477/pT479. We therefore next evaluated the contribution of Akt1-pS477/pT479 to Akt kinase activation under various cellular conditions.

In keeping with the notion that Akt1 tail phosphorylation is required to achieve full Akt1 activation, the phospho-deficient Akt1-S477A/T479A (Akt1-AA) mutant showed a dramatic suppression, whereas a phosphomimetic S477D/T479E (Akt1-DE) mutant showed enhanced Akt1 phosphorylation and activation (Fig. 3a–c and Supplementary Fig. 15a–c). However, caution needs to be taken in interpreting these results, as these mutations might not fully recapitulate the *in vivo* phosphorylation status. Strikingly, Akt1-AA was defective in pS473 even in the myristoylation-tagged constitutively active Akt1 (Fig. 3d), advocating a critical role for Akt1-pS477/pT479 in activating Akt (Supplementary Fig. 16a–c). Consistently, S477D/T479E effectively rescued Akt-pS473 and kinase activity of the cyclin A binding motif-defective mutant, R76A (Fig. 3e, f), as well as its anchorage-independent growth ability (Supplementary Fig. 16d). However, the rescue effect of DE on Akt1-R76A was not due to changes in its affinity with PIP3 (Supplementary Fig. 16e).

Consistent with S477/T479 being Cdk2/cyclin A sites, ectopic expression of Akt1-DE partly rescued cell cycle defects observed in quadruple knockout MEFs (cyclin E1<sup>-/-</sup>/cyclin E2<sup>-/-</sup>/cyclin A1<sup>-/-</sup>/cyclin A2<sup>trf</sup>) after Cre infection (Fig. 3g and Supplementary Fig. 16f). Compared with the periodic phosphorylation of wild-type Akt1 (Akt1-WT), Akt1-AA was severely impaired, whereas Akt1-DE showed an elevated and constitutive S473/T308 phosphorylation across the cell cycle (Fig. 3h). More importantly, under several physiological conditions (Supplementary Fig. 17a–c), Akt-pS473/pT308 was severely compromised in Akt1-AA, but robustly elevated in Akt1-DE expressing cells. Consistently, depletion of cyclin A2 in Akt1-WT, but not Akt1-DE expressing cells, led to a significant reduction in Akt phosphorylation (Supplementary Fig. 17d) and decreased tumour formation *in vivo* (Fig. 3i and Supplementary Fig. 17e–g). Cumulatively, these data demonstrate that in response to several upstream signals, phosphorylation of Akt1-S477/T479 may govern the canonical Akt-pS473 to promote Akt activation. Thus, we next



**Figure 2 | Cdk2/cyclin A2 functioned as a physiological kinase phosphorylating Akt1 at both S477 and T479.** **a**, Akt1 interacted with cyclin A2 at endogenous levels. Immunoblot analysis of WCLs and anti-Akt1 immunoprecipitations (IP) derived from MCF7 or T47D cells. **b**, Illustration of four putative cyclin A binding motifs (RXL) in human Akt1. **c**, Deficiency in cyclin A binding led to attenuated Akt phosphorylation. Immunoblot of WCLs and human influenza hemagglutinin (HA)-immunoprecipitations derived from HeLa cells transfected with indicated HA-Akt1 constructs. **d**, Immunoblot of WCLs derived from HeLa cells treated with indicated Cdk2 inhibitors at various time points. Drug doses used were roscovitine (40  $\mu$ M), CVT-131 (0.5  $\mu$ M) and mimosine (50  $\mu$ M). **e**, Immunoblot of WCLs derived from cyclin A2<sup>flf</sup> primary MEFs with or without adenoviral-Cre

infection. **f**, Immunoblot of WCLs derived from indicated immortalized MEFs with or without Cre viral infection. **g**, Immunoblot of WCLs derived from cyclin A2<sup>flf</sup> mouse embryonic stem (ES) cells stably expressing Cre-ER treated with 2  $\mu$ g ml<sup>-1</sup> tamoxifen (4-OHT) for the indicated days. **h**, Immunoblot of WCLs derived from HeLa cells stably expressing inducible pTRIPZ-cyclin-A2 treated with 500 ng ml<sup>-1</sup> doxycycline for the indicated periods. **i**, **j**, MDA-MB-231 cells depleted of Cdh1 by two independent shRNAs (**i**), or depleted of both Cdh1 and cyclin A2 (**j**), were injected into nude mice ( $n = 10$  for each group) and monitored for tumorigenesis *in vivo*. \* $P < 0.05$  (Student's *t*-test). **k**, Relative percentages of 50 samples examined from patients with breast cancer bearing the indicated cyclin A2 and Akt1-pS477/pT479 status indexed by either low or high.

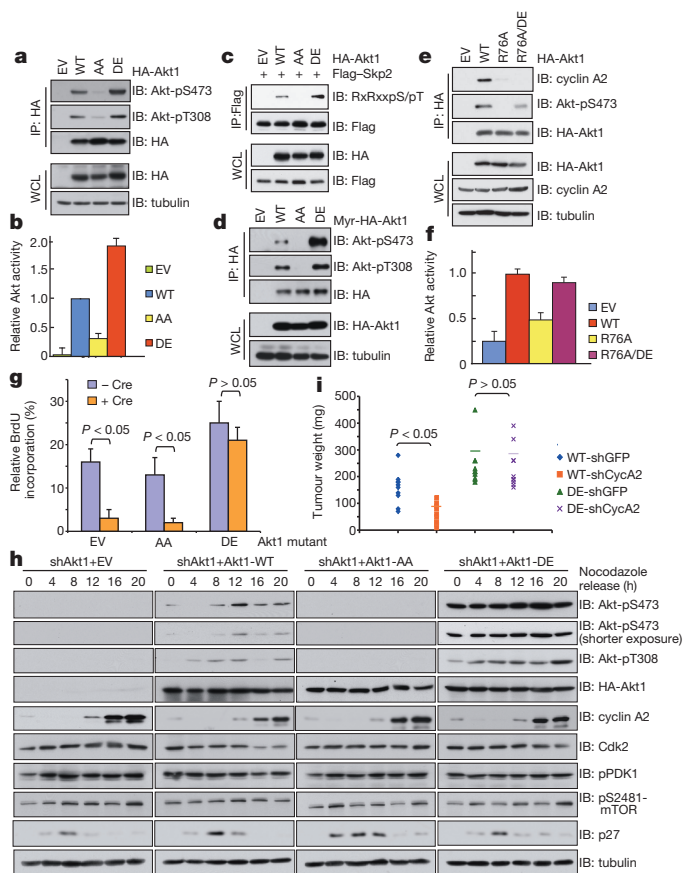
evaluated the precise molecular mechanism(s) linking Akt1-pS477/pT479 to pS473 and subsequent Akt activation.

As Akt2 crystal structures are reported in great detail<sup>22,23</sup>, we next focused on understanding how Akt2-pS478 (equivalent to Akt1-pS477; Supplementary Figs 4d and 18a) modulates Akt2 kinase activity. Notably, Akt2-pS478 functioned synergistically with Akt2-pS474 (equivalent to Akt1-pS473) to allosterically activate Akt2 (Fig. 4a and Supplementary Fig. 18b). Mechanistically, Akt2-pS478 may create a new charge-charge interaction between S478D and R208 to stabilize Akt2 in its closed, active form (Supplementary Fig. 18c, d). Consistently, deletion of the Akt1-tail after amino acid 475 (termed 476Δ) led to significantly increased Akt phosphorylation (Supplementary Fig. 18e), suggesting that either phosphorylation or deletion of the tail region could lock Akt in its active conformation.

Intriguingly, in addition to stabilizing Akt, Akt tail phosphorylation may alter Akt kinase kinetics to accelerate the kinase reaction (Supplementary Fig. 18f). Furthermore, pS473 and pS477/pT479, both of which activate Akt, are intrinsically linked; as compared with Akt1-WT, Akt1-DE but not Akt1-AA was preferentially phosphorylated by mTOR *in vitro* (Supplementary Fig. 19a, b). This indicates that pS477/pT479 might prime Akt1 for mTORC2-mediated phosphorylation of S473 (refs 5, 24) (Fig. 4b).

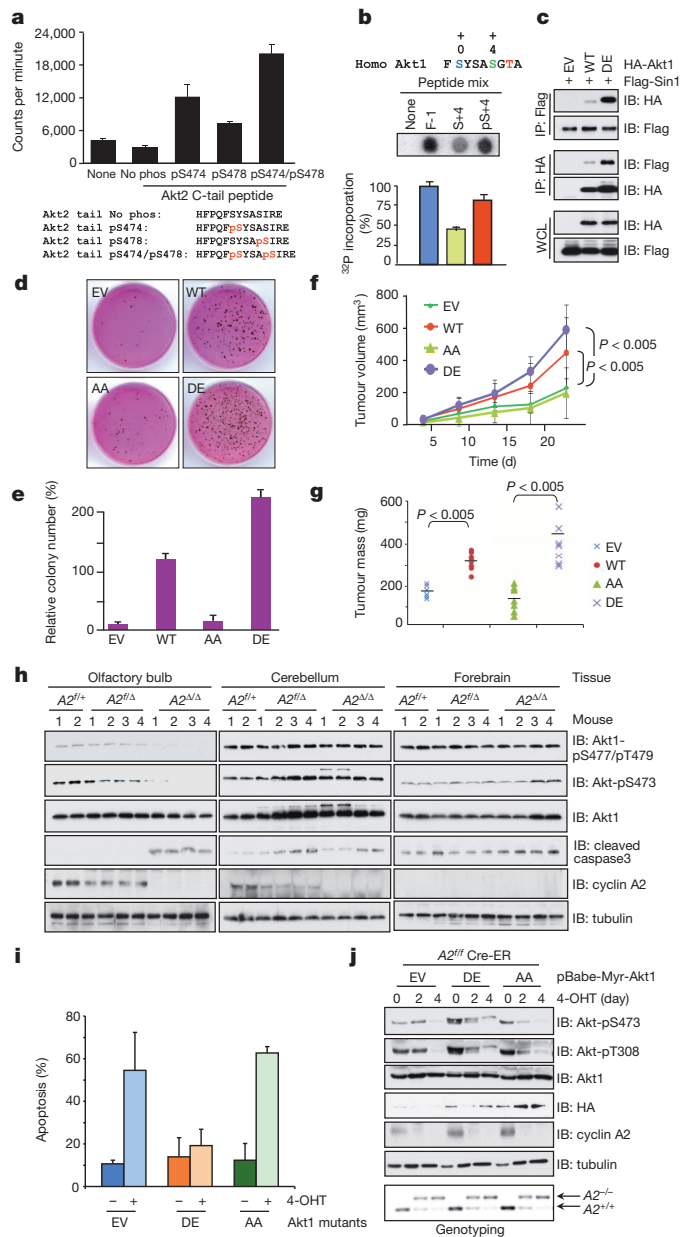
Mechanistically, enhanced phosphorylation of Akt1-DE at S473 by mTOR may be in part due to its increased interaction with Sin1 (Fig. 4c) and mTOR, but not due to changes in its association with the Akt inhibitor CTMP<sup>25</sup> or phosphatase PHLPP2 (ref. 26) (Supplementary Fig. 19c, d). However, phospho-mimetic mutation of S473 (S473D) (Supplementary Fig. 20a) or its mimicking peptide library (Supplementary Fig. 20b) led to reduced Akt1 tail phosphorylation, suggesting that the pS477/pT479 event may precede pS473.

Mutation of either S473 or S477/T479 to alanine significantly reduced Akt activity towards phosphorylating Skp2 or crosslides, whereas no significant additive effect was observed in S473A/S477A/T479A (Supplementary Fig. 21a, b), arguing for a possible functional redundancy between pS477/pT479 and pS473 for Akt catalytic activity. Consistently, Akt1-DE kinase activity was largely unchanged after depletion of Rictor (Supplementary Fig. 21c, d), whereas Akt1-DE partly rescued the deficient Akt kinase activity in Akt1-S473A (Supplementary Fig. 21e, f). Given the close proximity of S473 and S477/T479, these results indicate that in addition to promoting pS473, pS477/pT479 may also trigger Akt activation independently of, or partly compensate for, pS473.



**Figure 3 | Akt1-S477/T479 phosphorylation triggered Akt1-S473 phosphorylation and enhanced Akt1 activation.** **a**, Akt1-S477D/T479E mutation led to elevated Akt-pS473. Immunoblot of WCLs and HA-immunoprecipitations derived from HeLa cells transfected with indicated Akt1 constructs. EV, empty vector. **b**, *In vitro* kinase assays of indicated affinity-purified Akt1 kinases with crosside as a substrate. Experiments were performed in triplicate; data are shown as mean  $\pm$  s.d. **c**, Akt1-S477D/T479E mutation resulted in enhanced Skp2 phosphorylation. Immunoblot of WCLs and Flag-immunoprecipitations derived from HeLa cells transfected with indicated Akt1 constructs together with Flag-Skp2. **d**, **e**, Immunoblot of WCLs and HA-immunoprecipitations derived from HeLa cells transfected with indicated Akt1 constructs. **f**, *In vitro* kinase assays of indicated affinity-purified Akt1 kinases with crosside as a substrate. Experiments were performed in triplicate; data are shown as mean  $\pm$  s.d. **g**, Relative 5-bromodeoxyuridine (BrdU) incorporation for quadruple knockout MEFs (cyclin E1<sup>-/-</sup>/E2<sup>-/-</sup>/A1<sup>-/-</sup>/A2<sup>fl/fl</sup>) expressing indicated Akt1 constructs that were further infected with or without Cre to delete the cyclin A2 alleles. **h**, Akt1-S477D/T479E mutation led to sustained Akt phosphorylation across the cell cycle. Immunoblot of WCLs derived from HeLa cells transfected with indicated Akt1 constructs, synchronized by nocodazole and released for indicated periods. **i**, Akt1-depleted HeLa cells stably expressing Akt1-WT or -DE were subjected to cyclin A2 knockdown by lentiviral shRNA infections. The resulting cell lines were subcutaneously injected into nude mice ( $n = 10$  for each group) and monitored for tumorigenesis.  $P$  values were calculated by Student's  $t$ -test.

Biologically, phosphorylation of the Akt1 tail activated Akt1, leading to elevated Skp2 or FOXO phosphorylation, which further promoted cell cycle progression (Supplementary Fig. 22a–c), or conferred resistance to the chemotherapeutic agents etoposide or camptothecin (Supplementary Fig. 22d, e), respectively. More importantly, compared with Akt1-WT, Akt1-DE-expressing cells showed growth advantage in both *in vitro* soft agar (Fig. 4d, e) and *in vivo* tumour formation (Fig. 4f, g and Supplementary Fig. 23a, b) assays, supporting a role for Akt1-pS477/pT479 in promoting Akt1 activation and signalling phenotypes associated with malignancy. Notably, Akt1-476 $\Delta$  phenocopied Akt1-DE by showing elevated Akt phosphorylation (Supplementary Fig. 18e),



**Figure 4 | Akt tail phosphorylation triggered Akt activation to promote Akt oncogenic functions.** **a**, Stimulation of purified recombinant pAPH-Akt2- $\Delta$ C kinase activity by indicated Akt2 C-terminal tail peptides. pAPH-Akt2- $\Delta$ C, Akt2 residues 146–460, lacking the PH domain and C-terminal tail with T309 phosphorylated. Experiments were performed in triplicate; data are shown as mean  $\pm$  s.d. No phos, no phosphorylation. **b**, mTOR *in vitro* kinase assays with degenerate peptide libraries as substrates. Experiments were performed in triplicate; data are shown as mean  $\pm$  s.d. **c**, Akt1-S477D/T479E mutation led to enhanced binding with Sin1. Immunoblot of WCLs and HA- or Flag-immunoprecipitations derived from Akt1-depleted HeLa cells transfected with indicated Akt1 constructs together with Flag-Sin1. **d**, **e**, Soft agar assays using Akt1-depleted HeLa cells stably expressing WT-, AA- or DE-Akt1. **f**, **g**, Akt1-depleted HeLa cells stably expressing WT-, AA- or DE-Akt1 were injected subcutaneously into nude mice ( $n = 10$  for each group) and monitored for tumorigenesis (**f**). Tumours were dissected and weighed (**g**).  $*P < 0.05$  (Student's  $t$ -test). **h**, Immunoblot analyses of the indicated mouse brain tissues derived from mice with the indicated cyclin A2 genetic status. **i**, Fluorescence-activated cell sorting analysis of cyclin A2<sup>fl/fl</sup> mouse embryonic stem cells stably expressing Cre-ER in the presence of 2  $\mu$ g ml<sup>-1</sup> tamoxifen for 4 days. **j**, Immunoblot analysis of WCLs derived from various indicated embryonic stem stable cell lines generated in **i**.



probably through enhanced interaction with mTORC2 (Supplementary Fig. 24a), thereby promoting *in vitro* anchorage-independent growth and *in vivo* tumorigenesis (Supplementary Fig. 24 b–f).

To study the *in vivo* physiological significance of Cdk2/cyclin-A-mediated phosphorylation on Akt1-S477/T479 further, we generated brain-specific cyclin A2 knockout mice with nestin-Cre<sup>27</sup>. Interestingly, a significant reduction in Akt1-pS477/pT479 was observed in cyclin A2<sup>Δ/Δ</sup> olfactory bulbs coupled with elevated cleavage of caspase 3 (Fig. 4h), suggesting that Akt1-pS477/pT479 might be critical for cell survival in olfactory bulbs. Consistently, acute ablation of cyclin A2 in cyclin A2<sup>fl/fl</sup> mouse embryonic stem cells led to induced cellular apoptosis (Fig. 4i and Supplementary Fig. 25a, d), providing a possible explanation for previous findings that deleting cyclin A2 abolished embryonic stem cell colony formation *in vitro*<sup>14</sup>. Furthermore, cyclin A2-deletion-induced elevation of cellular apoptosis in mouse embryonic stem cells could be partly rescued by expressing Akt1-DE, but not Akt1-AA (Fig. 4i, j and Supplementary Fig. 25), supporting the idea that cyclin A2 may govern cellular survival *in vivo* largely by promoting Akt activation.

To extend these findings and their clinical relevance to human pathophysiology, we observed a positive correlation between Akt1-pS477/pT479 and Akt1-pS473 in samples from patients with breast cancer and breast-cancer-derived cell lines (Supplementary Fig. 26a–e). Interestingly, high levels of pS477/pT479 occurred at a relatively higher rate than pS473 in an earlier breast cancer developmental stage (stage II) (Supplementary Fig. 26d), indicating that pS477/pT479 may serve as a better biomarker for early-stage breast cancer detection, although further investigations are warranted.

Taken together, our data unravel a new phosphorylation event on Akt1 at its extreme C-terminal residues, S477 and T479, to trigger Akt1 activation either through enhancing the association between Akt1 and mTORC2 to promote pS473, or by functionally compensating for pS473 to lock Akt1 in its active conformation. More importantly, our study directly couples Akt activity with cell cycle progression, two well-characterized hallmarks of human cancers<sup>28–30</sup>. In this regard, our data suggest that cyclin A2 overexpression might exert its physiological functions in part by directly phosphorylating and activating Akt to trigger its pro-survival and oncogenic functions.

## METHODS SUMMARY

Expression plasmid constructs, proteins, antibodies, cell lines and the sequences of various short hairpin RNA (shRNA) oligonucleotides used in this study are described in Methods. All mutants were generated using mutagenesis PCR and the sequences were verified by DNA sequencing. Several pharmacological kinase inhibitors were used to screen for the kinases responsible for the Akt1-S477/T479 phosphorylation. The *in vivo* Akt1-S477/T479 phosphorylation events were identified by mass spectrometry and examined by *in vitro* kinase assays. The functional roles of Akt1-S477/T479 phosphorylation on Akt-S473 phosphorylation and kinase activity were examined either by co-immunoprecipitation or *in vitro* kinase assays. Cell viability assays were used to detect the Akt activity conferred by the Akt-S477/T479 phosphorylation status. Soft agar and mouse xenograft assays were performed to analyse physiologically how the Akt1-S477/T479 phosphorylation contributes to elevated oncogenic activity to facilitate anchorage-independent growth and *in vivo* tumour growth. Immunohistochemistry on clinical samples from patients with breast cancer was performed to illustrate a significant positive correlation between Akt-pS473 and Akt1-pS477/pT479. Brain-specific cyclin A2<sup>Δ/Δ</sup> mice were generated by crossing cyclin A2<sup>fl/fl</sup> mice with cyclin A2<sup>fl/fl</sup>-nestin-Cre mice, described in detail in Methods.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 June 2013; accepted 23 January 2014.

Published online 9 March 2014.

- Zoncu, R., Efeyan, A. & Sabatini, D. M. mTOR: from growth signal integration to cancer, diabetes and ageing. *Nature Rev. Mol. Cell Biol.* **12**, 21–35 (2010).
- Manning, B. D. & Cantley, L. C. AKT/PKB signaling: navigating downstream. *Cell* **129**, 1261–1274 (2007).
- Toker, A. Akt signaling: a damaging interaction makes good. *Trends Biochem. Sci.* **33**, 356–359 (2008).
- Luo, J., Manning, B. D. & Cantley, L. C. Targeting the PI3K-Akt pathway in human cancer: rationale and promise. *Cancer Cell* **4**, 257–262 (2003).
- Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).
- Stephens, L. et al. Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-trisphosphate-dependent activation of protein kinase B. *Science* **279**, 710–714 (1998).
- Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
- Gao, X. & Zhang, J. Spatiotemporal analysis of differential Akt regulation in plasma membrane microdomains. *Mol. Biol. Cell* **19**, 4366–4373 (2008).
- Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
- McGarry, T. J. & Kirschner, M. W. Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043–1053 (1998).
- Murphy, M. et al. Delayed early embryonic lethality following disruption of the murine cyclin A2 gene. *Nature Genet.* **15**, 83–86 (1997).
- Maddika, S. et al. Akt-mediated phosphorylation of CDK2 regulates its dual role in cell cycle progression and apoptosis. *J. Cell Sci.* **121**, 979–988 (2008).
- Adams, P. D. et al. Identification of a cyclin-cdk2 recognition motif present in substrates and p21-like cyclin-dependent kinase inhibitors. *Mol. Cell. Biol.* **16**, 6623–6633 (1996).
- Kalaszczynska, I. et al. Cyclin A is redundant in fibroblasts but essential in hematopoietic and embryonic stem cells. *Cell* **138**, 352–365 (2009).
- Geley, S. et al. Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J. Cell Biol.* **153**, 137–148 (2001).
- Malumbres, M. & Barbacid, M. Mammalian cyclin-dependent kinases. *Trends Biochem. Sci.* **30**, 630–641 (2005).
- Gray, C. H. & Barford, D. Getting in the ring: proline-directed substrate specificity in the cell cycle proteins Cdc14 and CDK2-cyclinA3. *Cell Cycle* **2**, 500–502 (2003).
- Liu, P., Kenney, J. M., Stiller, J. W. & Greenleaf, A. L. Genetic organization, length conservation, and evolution of RNA polymerase II carboxyl-terminal domain. *Mol. Biol. Evol.* **27**, 2628–2641 (2010).
- Stokes, M. P. et al. PTMScan direct: identification and quantification of peptides from critical signaling proteins by immunoaffinity enrichment coupled with LC-MS/MS. *Mol. Cell. Proteom.* **11**, 187–201 (2012).
- Bozulic, L., Surucu, B., Hynx, D. & Hemmings, B. A. PKB $\alpha$ /Akt1 acts downstream of DNA-PK in the DNA double-strand break response and promotes survival. *Mol. Cell* **30**, 203–213 (2008).
- Xie, X. et al. I $\kappa$ B kinase epsilon and TANK-binding kinase 1 activate AKT by direct phosphorylation. *Proc. Natl Acad. Sci. USA* **108**, 6474–6479 (2011).
- Yang, J. et al. Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. *Nature Struct. Biol.* **9**, 940–944 (2002).
- Yang, J. et al. Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. *Mol. Cell* **9**, 1227–1240 (2002).
- Alessi, D. R., Pearce, L. R. & Garcia-Martinez, J. M. New insights into mTOR signaling: mTORC2 and beyond. *Sci. Signal.* **2**, pe27 (2009).
- Maira, S. M. et al. Carboxyl-terminal modulator protein (CTMP), a negative regulator of PKB/Akt and v-Akt at the plasma membrane. *Science* **294**, 374–380 (2001).
- Gao, T., Furnari, F. & Newton, A. C. PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth. *Mol. Cell* **18**, 13–24 (2005).
- Gaveriaux-Ruff, C. & Kieffer, B. L. Conditional gene targeting in the mouse nervous system: Insights into brain function and diseases. *Pharmacol. Ther.* **113**, 619–634 (2007).
- Burgess, D. J. Senescence: double or quit? *Nature Rev. Cancer* **11**, 389 (2011).
- Butt, A. J. et al. Cell cycle machinery: links with genesis and treatment of breast cancer. *Adv. Exp. Med. Biol.* **630**, 189–205 (2008).
- Altomare, D. A. & Testa, J. R. Perturbations of the AKT signaling pathway in human cancer. *Oncogene* **24**, 7455–7464 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank J. Guo, J.J. Liu, A.W. Lau, S. Shaik, A. Tron, X. Dai and K. Xu for reading the manuscript, S.B. Breitkopf for help with mass spectrometry experiments, Y. Geng, L. Liu, K. Ran, R. Chin and S. Elloul for providing reagents, and members of the Wei, Tokar, Sicinski, Pandolfi and Cantley laboratories for discussions. W.W. is an American Cancer Society and a Leukemia & Lymphoma Society research scholar. P.L. is supported by 5T32HL007893. This work was supported in part by National Institutes of Health grants to W.W. (GM089763, GM094777 and CA177910), J.M.A. (2P01CA120964) and P.S. (R01CA132740).

**Author Contributions** P.L., M.B., W.M., H.I., A.P., M.G., D.G., P.T. and W.G. performed most of the experiments with assistance from B.K., L.W., A.S., B.Z. and M.Y. W.W., P.S., P.P.P., L.C. and P.L. designed the experiments. W.W., L.C., P.S., P.P.P., M.W.K. and A.T. supervised the study. P.L. and W.W. wrote the manuscript. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.W. ([wwwei2@bidmc.harvard.edu](mailto:wwwei2@bidmc.harvard.edu)).

# Nectar secretion requires sucrose phosphate synthases and the sugar transporter SWEET9

I Winnie Lin<sup>1,2</sup>, Davide Sosso<sup>1,2</sup>, Li-Qing Chen<sup>2</sup>, Klaus Gase<sup>3</sup>, Sang-Gyu Kim<sup>3</sup>, Danny Kessler<sup>3</sup>, Peter M. Klinkenberg<sup>4†</sup>, Molly K. Gorder<sup>4†</sup>, Bi-Huei Hou<sup>2</sup>, Xiao-Qing Qu<sup>2,5</sup>, Clay J. Carter<sup>4†</sup>, Ian T. Baldwin<sup>3</sup> & Wolf B. Frommer<sup>1,2</sup>

**Angiosperms developed floral nectaries that reward pollinating insects<sup>1</sup>. Although nectar function and composition have been characterized, the mechanism of nectar secretion has remained unclear<sup>2</sup>. Here we identify SWEET9 as a nectary-specific sugar transporter in three eudicot species: *Arabidopsis thaliana*, *Brassica rapa* (extrastaminal nectaries) and *Nicotiana attenuata* (gynoecial nectaries). We show that SWEET9 is essential for nectar production and can function as an efflux transporter. We also show that sucrose phosphate synthase genes, encoding key enzymes for sucrose biosynthesis, are highly expressed in nectaries and that their expression is also essential for nectar secretion. Together these data are consistent with a model in which sucrose is synthesized in the nectary parenchyma and subsequently secreted into the extracellular space via SWEET9, where sucrose is hydrolysed by an apoplasmic invertase to produce a mixture of sucrose, glucose and fructose. The recruitment of SWEET9 for sucrose export may have been a key innovation, and could have coincided with the evolution of core eudicots and contributed to the evolution of nectar secretion to reward pollinators.**

Plants have evolved anatomical and physiological features for attracting animals to promote pollination. Reproductive isolation and thus speciation is thought to be enhanced in animal-pollinated species relative to those reliant on wind pollination<sup>3</sup>. Floral traits, including animal pollination, can affect subsequent species abundance within clades<sup>1</sup>. When de Saporta, Hooker, Heer and Darwin discussed the “abominable mystery”—the apparent rapid radiation of angiosperms and insects in the mid-Cretaceous period, de Saporta suggested that the development and refinement of insect-assisted pollination through the coevolution of pollinators and flowering plants may have been key to pollinator and angiosperm diversification<sup>4</sup>. Yet the molecular mechanism of nectar secretion has remained elusive<sup>2</sup>.

Flowering plants evolved intricate methods to secure efficient interaction with pollinators and, thereby, successful reproduction and genetic diversity through cross-pollination. Nectar, which contains sugars and volatile compounds that attract and reward pollinators<sup>5</sup> and toxins that repel unwanted visitors<sup>6</sup>, compels pollinators to optimize outcrossing rates<sup>7</sup>. Nectar composition varies between species of plants, possibly to reward different animal species<sup>2</sup>. Depending on the plant species, sucrose, glucose and fructose make up 8–80% (w/w) of nectar. Angiosperm nectar is synthesized and secreted by different types of floral nectaries (FNs) and extrafloral nectaries (EFNs). *N. attenuata* (coyote tobacco), a self-compatible, hawkmoth- and hummingbird-pollinated asterid, produces nectar containing sucrose, hexoses and numerous secondary metabolites<sup>6,8</sup>. *B. rapa* (turnip), comprising self-compatible and incompatible varieties, produces hexose-dominant nectar<sup>9,10</sup>. *A. thaliana*, a self-compatible self-fertilizer, develops functional nectaries that produce volatiles and hexose-rich nectar<sup>9,11,12</sup>. Whether nectar production in self-fertilizing plants represents an evolutionary remnant or functions to secure the observed low outcrossing rates has yet to be established<sup>13</sup>.

To identify transporters involved in nectar secretion, we searched for candidate sugar transporters specifically expressed in nectaries. SWEET sugar transporters appeared as prime candidates for a role in nectar secretion<sup>14,15</sup>: for example, *A. thaliana* (At)SWEET11 and 12 are responsible for sucrose efflux, a key step in sucrose translocation from photosynthetic tissues to heterotrophic tissues. A petunia SWEET9 homologue (PhNEC1; Extended Data Figs 1, 2) had been found in nectaries, and expression correlated inversely with nectary starch content<sup>16</sup>; however, its role remained elusive until the *Arabidopsis* and rice homologues were shown to function as sugar transporters<sup>15</sup>. Here we show the critical role of SWEET9 in nectar secretion by confirming its expression in nectaries, sucrose transport activity, and localization at the plasma membrane. Mutation of AtSWEET9 or nectary-expressed sucrose phosphate synthase (SPS) genes led to the loss of nectar secretion. Sugars delivered to defective nectaries accumulated in the stems at the floral base, indicating the lack of negative feedback on phloem delivery and the inability to relocate incoming sucrose. We demonstrate that SWEET9 is either conserved or has been independently co-opted in rosids (*Arabidopsis*, turnip) and an asterid (coyote tobacco) for nectar secretion.

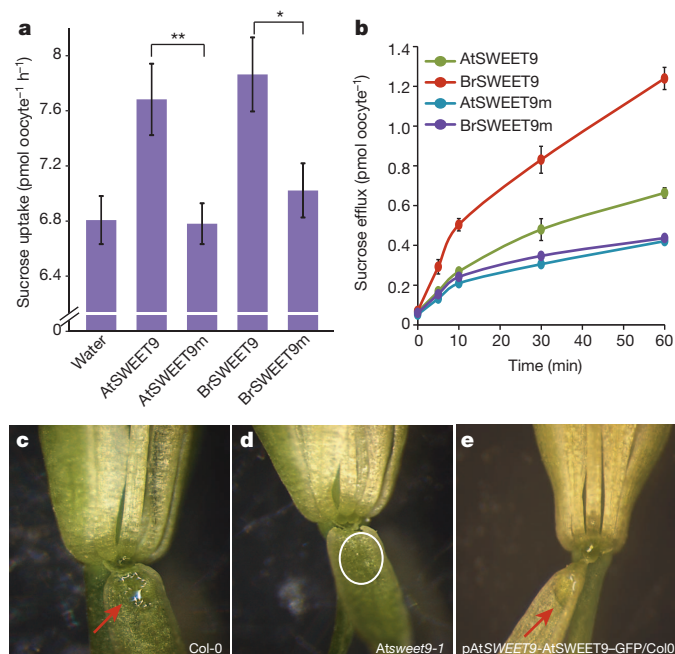
AtSWEET9, which shares ~50% identity with AtSWEET11 or 12 (clade 3 of the SWEET family; Extended Data Figs 1, 2), is highly expressed in *Arabidopsis* nectaries<sup>17</sup> (Extended Data Fig. 3). Transport studies in *Xenopus* oocytes showed that AtSWEET9 mediates sucrose uptake and efflux (Fig. 1a, b). The sucrose transport activity of AtSWEET9 was confirmed by coexpression of AtSWEET9 with fluorescence resonance energy transfer (FRET) sucrose sensors in human embryonic kidney cells<sup>14</sup>, indicating that AtSWEET9 functions as a facilitated diffusion transporter for sucrose<sup>15</sup>. AtSWEET9 weakly transports glucose; possibly contributing to hexose efflux (Extended Data Fig. 4).

To determine directly whether AtSWEET9 is involved in nectar secretion, we examined nectar in transfer DNA (T-DNA) insertion mutants. Depending on whether AtSWEET9 has a role in sugar uptake into or efflux from nectaries, nectar from mutants would either contain reduced sugar content or lose fluid secretion. *Atsweet9* contained no detectable nectar droplets (Fig. 1c, d, Extended Data Fig. 5 and Supplementary Table 1). To test whether AtSWEET9 activity limits nectar secretion, extra AtSWEET9 copies expressed from the native promoter in wild-type plants increased nectar volume and glucose amount (Fig. 1e and Supplementary Table 2). Restoration of nectar secretion by AtSWEET9 or AtSWEET9 tagged with green fluorescent protein (GFP) in *Atsweet9* mutants further supports the role of SWEET9 in nectar secretion (Extended Data Fig. 5).

SWEET9 could function in (1) sucrose efflux from phloem, (2) sugar uptake into, or (3) sugar efflux from nectary parenchyma. Translational  $\beta$ -glucuronidase (GUS) and enhanced (e)GFP fusions driven by the AtSWEET9 promoter were specifically expressed in nectaries (Fig. 2a–c), with the highest levels seen in the lower half of nectary parenchyma, but not in guard cells and phloem. The fluorescence intensity of

<sup>1</sup>Department of Biology, Stanford University, Stanford, California 94305, USA. <sup>2</sup>Carnegie Institution for Science, 260 Panama Street, Stanford, California 94305, USA. <sup>3</sup>Max Planck Institute for Chemical Ecology, Jena D-07745, Germany. <sup>4</sup>Department of Biology, University of Minnesota Duluth, Duluth, Minnesota 55812, USA. <sup>5</sup>Key Laboratory of Plant and Soil Interactions, College of Resources and Environmental Sciences, China Agricultural University, 100193 Beijing, China. <sup>†</sup>Present address: Department of Plant Biology, University of Minnesota, St Paul, Minnesota 55108, USA.

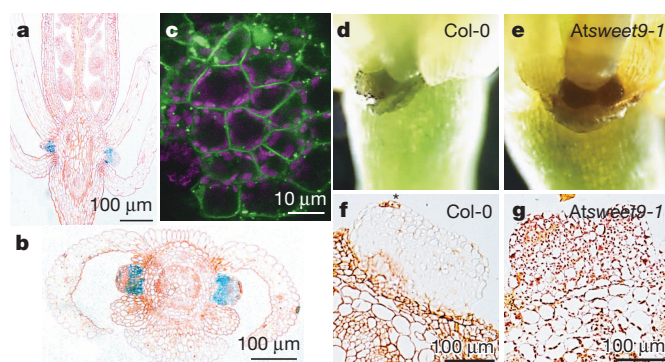




**Figure 1 | AtSWEET9, a sucrose transporter, is necessary for nectar secretion.** **a**, **b**, Sucrose uptake (**a**) and efflux (**b**) activity of AtSWEET9 and BrSWEET9 were determined in *Xenopus* oocytes. Truncated AtSWEET9(F201\*) (AtSWEET9m) and BrSWEET9(L201\*) (BrSWEET9m) served as negative controls. **a**, Oocyte uptake assay: AtSWEET9 and BrSWEET9 mediate <sup>14</sup>C-sucrose uptake (error bars show  $\pm$  standard error of the mean (s.e.m.),  $n = 14$ ,  $n = 16$ ). \* $P < 0.05$ , \*\* $P < 0.01$ , Student's *t*-test. **b**, Oocyte efflux assay: <sup>14</sup>C-sucrose efflux by AtSWEET9 and BrSWEET9 in *Xenopus* oocytes injected with <sup>14</sup>C-sucrose (error bars show  $\pm$  s.e.m.,  $n = 9$ ,  $n = 8$ ). **c**, Nectar droplet clinging to inside of sepal (wild type: Col-0). **d**, Lack of nectar in nectaries of *Atsweet9-1* mutants. **e**, Increased nectar secreted from nectaries of flowers containing extra copies of AtSWEET9-eGFP, driven by the AtSWEET9 promoter (pAtSWEET9). **c**–**e**, Original magnification,  $\times 10$ .

AtSWEET9-eGFP in floral nectaries increased during maturation, becoming highest during maximal nectar secretion (Extended Data Fig. 6). These data do not support a role for AtSWEET9 in phloem unloading, but do not enable us to distinguish between roles in uptake versus efflux from nectarial parenchyma. We conjectured that starch accumulation in *Atsweet9* nectaries would be different if SWEET9 were involved in uptake (no starch accumulation in nectary) versus cellular efflux from nectarial cells (accumulation in nectary due to export deficiency). In wild-type plants, starch accumulates in plastids of nectary parenchyma before anthesis and is degraded to produce sugars for secretion<sup>16,18</sup>. Mutants accumulated starch in all cells of the nectary parenchyma, indicating that AtSWEET9 is responsible for cellular sugar efflux (Fig. 2d–g). Interestingly, at anthesis, nectarial guard cells of wild-type, but not *Atsweet9* plants, contained starch granules, possibly due to reabsorption of sugars from nectar<sup>19</sup>. The functional characterization of AtSWEET9 as a sucrose efflux transporter, the expression of SWEET9 in nectary parenchyma and the starch accumulation pattern in mutants intimate AtSWEET9 as the key transporter for nectarial sugar secretion. The involvement of a putative facilitated-diffusion carrier indicates that high cytosolic levels of sugars in the nectarial parenchyma together with extracellular hydrolysis of sucrose by apoplasmic invertase create the driving force for nectar secretion<sup>9</sup>.

Whether sugar secretion is mediated by plasma membrane carriers or exocytosis has been a matter of debate<sup>20</sup>. AtSWEET9-eGFP fusions localized at the plasma membrane, the *trans*-Golgi apparatus and multivesicular bodies (Fig. 2c, Extended Data Fig. 7 and Supplementary Videos 1–3). Given the strength of the SWEET9 promoter, accumulation in the *trans*-Golgi/multivesicular bodies is probably caused by high



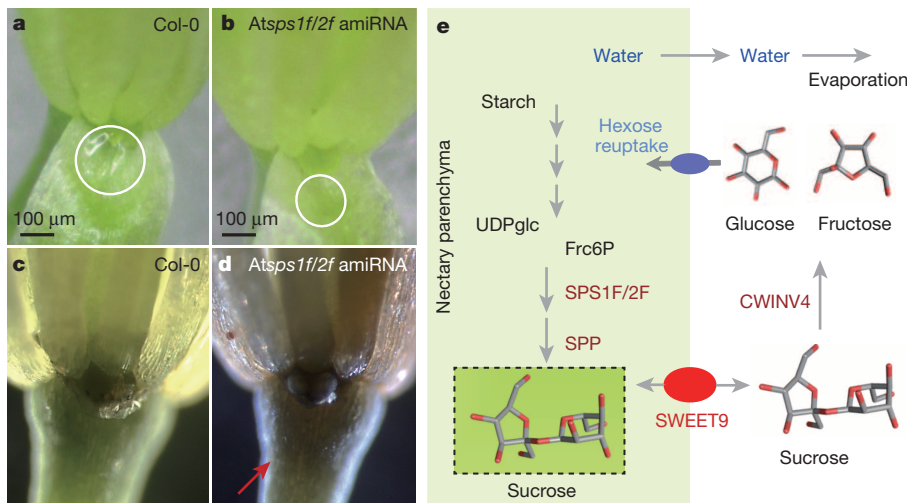
**Figure 2 | Cellular and subcellular localization of AtSWEET9 and starch accumulation in *Atsweet9* mutants.** **a**, **b**, Histochemical GUS analysis in *Arabidopsis* flowers expressing translational GUS fusion of AtSWEET9 (native promoter). Vertical (**a**) and transverse (**b**) sections of *Arabidopsis* flowers show tissue-specific localization of AtSWEET9. Cell walls stained with safranin-O (orange). **c**, Confocal images of eGFP fluorescence of an AtSWEET9 promoter:AtSWEET9-eGFP fusion showing subcellular localization at the plasma membrane and *trans*-Golgi network. **d**, **e**, Close-up of nectaries from wild-type and *Atsweet9-1* flowers stained with Lugol's iodine solution. Starch accumulated only in the guard cells of wild-type nectaries and in nectary parenchyma in *Atsweet9-1* (sampled at the end of the dark period). **f**, **g**, LR White resin sections of *Arabidopsis* nectaries in wild type and *Atsweet9-1* mutants stained with Lugol's iodine solution. Starch grains (dark red) accumulate in nectaries of *Atsweet9-1* mutants (**g**) and in stomata of wild-type nectary (**f**, indicated by an asterisk). Starch grains in floral stalks and nectaries in wild-type and *Atsweet9* mutant lines at anthesis. Cell walls stained with safranin-O (orange).

expression levels<sup>21</sup>. Thus, although it is conceivable that AtSWEET9 loads vesicles with sugars for exocytosis, plasma-membrane-mediated efflux is more likely. Given that multiple clade 3 SWEETs transport sucrose, we explored whether the well-characterized plasma-membrane-localized paralogues AtSWEET11 and 12 (ref. 15) can complement the loss of nectar secretion in *Atsweet9*. When expressed from the AtSWEET9 promoter, both transporters restored nectar secretion (Extended Data Figs 5g, h, 6j–k and Supplementary Video 4), indicating that a sugar efflux transporter driven by a promoter that expresses it in nectaries suffices for sugar efflux.

Gene expression analysis in *Arabidopsis* has shown that sucrose biosynthesis genes are upregulated in nectaries<sup>17</sup>, intimating that resynthesis of sucrose derived from starch fuels sugar efflux. We found that two SPS genes, *SPS1F* and *SPS2F*, which encode key enzymes for sucrose biosynthesis, were highly expressed in maturing nectaries (Extended Data Fig. 8). Artificial microRNA (amiRNA) inhibition of SPS gene expression led to loss of nectar secretion and increased starch accumulation (Fig. 3a–d). Thus, the phenotypes of *Atsweet9* mutants, *SPS* amiRNA lines and the nectarial apoplasmic invertase mutant *Atcwin4-1* (ref. 22) are similar. The most parsimonious hypothesis is thus that starch-derived sucrose synthesized in nectaries is exported by SWEET9, and that sucrose hydrolysis by CWINV4 is necessary to create a large enough osmotic gradient to sustain water secretion (Fig. 3e). Species secreting nectar with a very high sucrose content may use additional mechanisms for creating a sufficient osmotic gradient.

To explore whether SWEET9 of other Brassicaceae is essential for sugar efflux from nectaries, we identified the orthologous *B. rapa* (Br)SWEET9 in turnip flowers and showed that it transports sucrose (Fig. 1a, b) and is essential for sugar efflux and nectar secretion (Fig. 4a, b and Extended Data Fig. 9a). Nectar from the rosids *Arabidopsis* and turnip contains predominantly hexoses<sup>9</sup>, consistent with the role of apoplasmic invertase in post-secretory sucrose hydrolysis. Other rosids may also use SWEET9 for nectar secretion; for example, microarray data indicate that SWEET9 has a role in nectar secretion from extrastaminal nectaries as it is the highest expressed SWEET in barrel clover (*Medicago truncatula*) flowers (Extended Data Fig. 10).

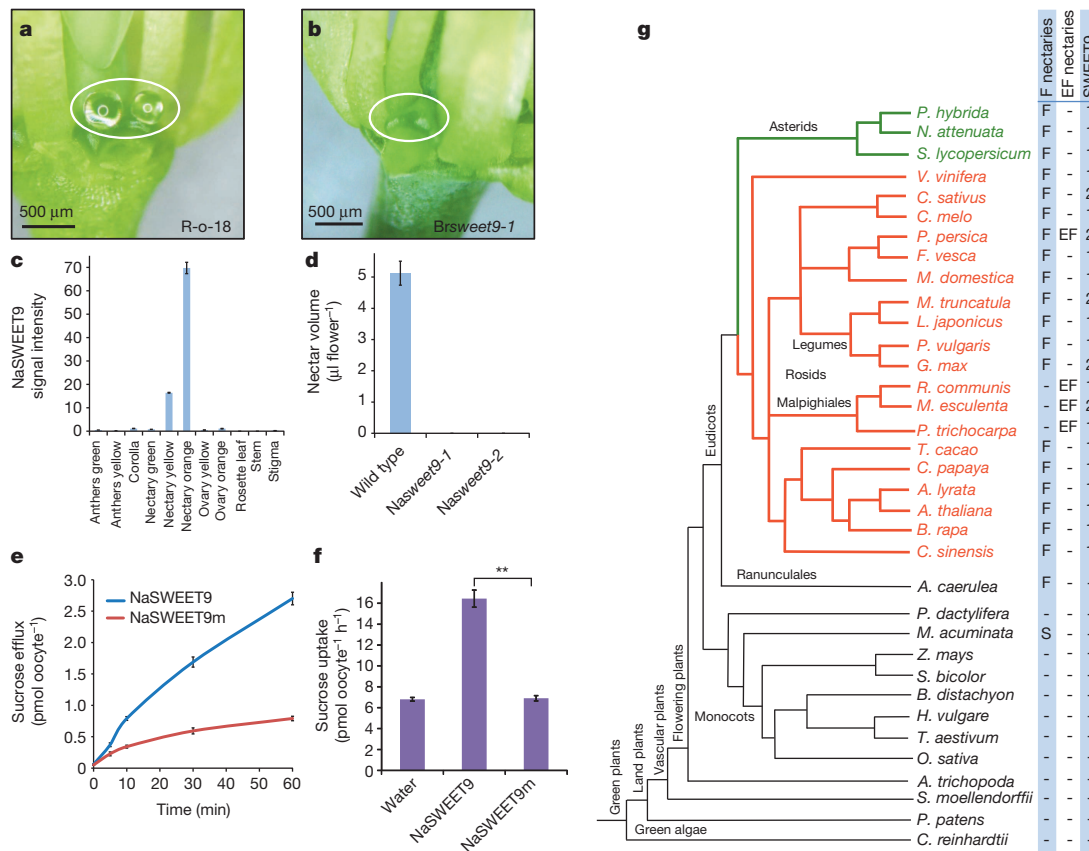




**Figure 3 | SPS1 and SPS2 are necessary for nectar secretion in *Arabidopsis*.** **a, b**, amiRNA inhibition of AtSPS1 and AtSPS2 gene expression led to a loss of nectar secretion. The circles indicate nectar droplets in wild-type flowers (**a**), and the absence of nectar in mutant flowers (**b**). **c, d**, Starch accumulation in nectaries of *Atsps* miRNA lines (floral stalk of *sps1f/2f* mutant lines, red arrow; only in guard cells of wild-type nectaries). Original magnification,  $\times 10$ . **e**, Model for nectar secretion: starch-derived sucrose is synthesized in nectaries by SPS proteins and exported by SWEET9. Exported sucrose is subsequently hydrolysed by CWINV4, which creates a high osmotic potential to sustain water flow down the osmotic gradient. Frc6P, fructose 6-phosphate; UDPglc, uracil diphosphate glucose.

Nectaries probably evolved multiple times<sup>23</sup>. Solanales (asterids) have gynoecial nectaries, whereas Brassicales typically have extra- and intra-staminal nectaries. To identify potential functional orthologues from coyote tobacco, *SWEET*s were cloned using degenerate primers (Supplementary Table 3). *N. attenuata* (Na)SWEET9 was most highly expressed in gynoecial nectaries; levels increased during nectary maturation (Fig. 4c). NaSWEET9 mediated sucrose uptake and efflux when expressed in oocytes

(Fig. 4e, f). NaSWEET9 was essential for nectar secretion as shown in two independent RNA interference (RNAi) lines (Fig. 4d and Extended Data Fig. 9). Together, our results indicate that SWEET9 either arose early during core eudicot evolution or has been co-opted independently in asterids and rosids for gynoecial and both intra- and extrastaminal nectary function. A phylogenetic analysis tentatively traces the appearance of SWEET9 to a time window nearly at the origin of core eudicots



**Figure 4 | SWEET9 orthologues in *B. rapa* and *N. attenuata* are essential for nectar secretion.** **a**, Nectar droplets in lateral nectary of wild-type (R-o-18) *B. rapa* flowers. **b**, Lack of nectar in Brsweet9-1 mutants. **c**, NaSWEET9 transcript accumulation in *N. attenuata* (error bars show  $\pm$  s.e.m.,  $n = 3$ ). **d**, Mean ( $\pm$  s.e.m.,  $n = 8$ ) nectar volume of flowers measured at 05:00 in wild-type, Nasweet9-1 and Nasweet9-2 plants. **e, f**, Sucrose uptake (**e**) and efflux (**f**) activity of NaSWEET9 in oocytes. Truncated version of NaSWEET9 (L201\*) (NaSWEET9m) served as control. **e**, Oocyte uptake: NaSWEET9 mediates  $^{14}\text{C}$ -sucrose uptake (error bars show  $\pm$  s.e.m.,  $n = 14$ ). **f**,  $^{14}\text{C}$ -sucrose efflux by

NaSWEET9 in oocytes (error bars show  $\pm$  s.e.m.,  $n = 10$ ).  $**P < 0.01$ , Student's *t*-test. **g**, Data were collected from genome databases (<http://phytozome.org>; <http://genomeevolution.org>; <http://bioinformatics.psb.ugent.be/plaza/>) using AtSWEET9 as bait. Tree branches are a schematic representation (Angiosperm Phylogeny Group III system). Species belonging to the core eudicot clades of rosids or asterids are underlined in red and green, respectively; full names are listed in Supplementary Table 5. Species bearing floral (F), extrafloral (EF) and sepal (S) nectaries are indicated.

~120 million years ago<sup>24</sup> (Extended Data Figs 1, 2). In all genomes analysed, including grasses, *Selaginella* and *Physcomitrella*, we found multiple SWEET9 paralogs. Thus SWEET9 may have evolved from other clade 3 members when core eudicots merged. SWEET9 seems to be absent from wind-pollinated rice and maize lacking nectaries, banana, which has a typical monocot septal nectary, and *Amborella*, a member of an early angiosperm lineage. *Aquilegia* is a highly derived genus within the Ranunculales at the base of the eudicots<sup>25</sup> and relies on entomophilous pollination mediated by elaborate nectar-filled nectar spurs. However, its genome seems to lack SWEET9 orthologues (Phytozome search, October 2013). The hypothesis of a single origin of SWEET9 during the evolution of core eudicots is consistent with conclusions from analyses of the master regulator of nectary development *CRABS CLAW*<sup>26</sup>.

Castor bean and cassava, which produce nectar from EFNs, also contain SWEET9 homologues<sup>27</sup> (Fig. 4g and Extended Data Figs 1, 2). We found that the wind-pollinated poplar, which lacks functional EFNs but has foliar EFNs, contains SWEET9 and SWEET10 homologues. The sucrose transporter *Populus trichocarpa* (Pt)SWEET10a (Extended Data Fig. 4) is highly expressed in EFNs<sup>28</sup>, indicating that the foliar EFN uses SWEET9 paralogs for nectar secretion. Poplar nectar is hexose dominant, potentially due to cell wall invertases, which are highly expressed in the EFN<sup>28</sup>. These EFNs probably have a role in attracting ants as a defence against other insects<sup>29</sup>.

Starch accumulation in floral stalks of the mutants indicates that phloem-derived sucrose is imported into nectaries symplasmically. Sucrose is then hydrolysed and stored as hexoses in the vacuole, or as starch<sup>18</sup>. During nectary maturation, sucrose is resynthesized via SPS, and SWEET9 begins to export sucrose down a concentration gradient, leading to sucrose accumulation in the apoplast (Fig. 3e). As SWEET9 seems to function as a uniporter, and because the cytosol contains other solutes that contribute to the osmotic potential, uniporter-driven efflux is insufficient for osmotically driven water secretion. Thus, apoplasmic sucrose has to be hydrolysed by secreted invertases to produce glucose and fructose, increasing the osmotic driving force and allowing water to be secreted. Ultimately, sugary nectar is secreted through open stomata. Together, our results show that SWEET9 serves as a sugar efflux transporter of the nectary parenchyma and is necessary for the secretion of nectar in two types of eudicot nectaries. Microarray data show that hexose-transporting sugar transporters (STPs) are expressed in nectaries<sup>12</sup>, indicating that they have a role in reuptake. The relative activities of apoplasmic invertase combined with selective reuptake possibly determine the final ratio of sucrose, fructose and glucose.

## METHODS SUMMARY

Heterologous expression of SWEET9 from in HEK293T cells and *Xenopus* oocytes was performed as described<sup>15</sup>. Mutant *Arabidopsis* was obtained from the *Arabidopsis* Biological Research Center (ABRC), insertion sites and transcript levels were verified by PCR and qRT-PCR, respectively. BrSWEET9 TILLING mutants were obtained from RevGen UK. NaSWEET9 was silenced in *N. attenuata* by RNAi (inverted repeat constructs). *SPS1F* and *SPS2F* were co-silenced through a single amiRNA targeting the messenger RNAs for both genes. Nectar secretion was evaluated microscopically. Starch was stained using potassium iodide<sup>22</sup>.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 12 June 2013; accepted 27 January 2014.**

**Published online 16 March; corrected online 23 April 2014 (see full-text HTML version for details).**

- Sargent, R. D. Floral symmetry affects speciation rates in angiosperms. *Proc. R. Soc. Lond. B* **271**, 603–608 (2004).
- Nicolson, S. W., Nepi, M. & Pacini, E. *Nectaries and Nectar* (Springer, 2007).
- Kay, K. M. et al. *Floral Characters and Species Diversification* (Oxford Univ. Press, 2006).
- Friedman, W. E. The meaning of Darwin's 'abominable mystery'. *Am. J. Bot.* **96**, 5–21 (2009).
- De la Barrera, E. & Nobel, P. S. Nectar: properties, floral aspects, and speculations on origin. *Trends Plant Sci.* **9**, 65–69 (2004).
- Kessler, D., Gase, K. & Baldwin, I. T. Field experiments with transformed plants reveal the sense of floral scents. *Science* **321**, 1200–1202 (2008).
- Kessler, D. et al. Unpredictability of nectar nicotine promotes outcrossing by hummingbirds in *Nicotiana attenuata*. *Plant J.* **71**, 529–538 (2012).
- Sime, K. R. & Baldwin, I. T. Opportunistic out-crossing in *Nicotiana attenuata* (Solanaceae), a predominantly self-fertilizing native tobacco. *BMC Ecol.* **3**, 6 (2003).
- Davis, A. R., Pylatuk, J. D., Paradis, J. C. & Low, N. H. Nectar-carbohydrate production and composition vary in relation to nectary anatomy and location within individual flowers of several species of Brassicaceae. *Planta* **205**, 305–318 (1998).
- Isokawa, S. et al. Novel self-compatible lines of *Brassica rapa* L. isolated from the Japanese bulk-populations. *Genes Genet. Syst.* **85**, 87–96 (2010).
- Huang, M. et al. The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (E)- $\beta$ -caryophyllene, is a defense against a bacterial pathogen. *New Phytol.* **193**, 997–1008 (2012).
- Kram, B. W. & Carter, C. J. *Arabidopsis thaliana* as a model for functional nectary analysis. *Sex. Plant Reprod.* **22**, 235–246 (2009).
- Hoffmann, M. H. et al. Flower visitors in a natural population of *Arabidopsis thaliana*. *Plant Biol.* **5**, 491–494 (2003).
- Chen, L. Q. et al. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).
- Chen, L. Q. et al. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. *Science* **335**, 207–211 (2012).
- Ge, Y. X. et al. NEC1, a novel gene, highly expressed in nectary tissue of *Petunia hybrida*. *Plant J.* **24**, 725–734 (2000).
- Kram, B. W., Xu, W. W. & Carter, C. J. Uncovering the *Arabidopsis thaliana* nectary transcriptome: investigation of differential gene expression in floral nectariferous tissues. *BMC Plant Biol.* **9**, 92 (2009).
- Ren, G. et al. Transient starch metabolism in ornamental tobacco floral nectaries regulates nectar composition and release. *Plant Sci.* **173**, 277–290 (2007).
- Langenberger, M. W. & Davis, A. R. Temporal changes in floral nectar production, reabsorption, and composition associated with dichogamy in annual caraway (*Carum carvi*; Apiaceae). *Am. J. Bot.* **89**, 1588–1598 (2002).
- Fahn, A. Structure and function of secretory cells. *Adv. Bot. Res.* **31**, 37–75 (2000).
- Gomord, V. et al. The C-terminal HDEL sequence is sufficient for retention of secretory proteins in the endoplasmic reticulum (ER) but promotes vacuolar targeting of proteins that escape the ER. *Plant J.* **11**, 313–325 (1997).
- Ruhlmann, J. M., Kram, B. W. & Carter, C. J. CELL WALL INVERTASE 4 is required for nectar production in *Arabidopsis*. *J. Exp. Bot.* **61**, 395–404 (2010).
- Johnson, S. D., Hobbhahn, N. & Bytebier, B. Ancestral deceit and labile evolution of nectar production in the African orchid genus *Disa*. *Biol. Lett.* **9**, 20130500 (2013).
- Bell, C. D., Soltis, D. E. & Soltis, P. S. The age and diversification of the angiosperms re-visited. *Am. J. Bot.* **97**, 1296–1303 (2010).
- Fior, S. et al. Spatiotemporal reconstruction of the *Aquilegia* rapid radiation through next-generation sequencing of rapidly evolving cpDNA regions. *New Phytol.* **198**, 579–592 (2013).
- Lee, J. Y. et al. Recruitment of CRABS CLAW to promote nectary development within the eudicot clade. *Development* **132**, 5021–5032 (2005).
- Rizzardo, R. A., Milfont, M. O., Silva, E. M. & Freitas, B. M. *Apis mellifera* pollination improves agronomic productivity of anemophilous castor bean (*Ricinus communis*). *An. Acad. Bras. Cienc.* **84**, 1137–1145 (2012).
- Escalante-Pérez, M. et al. Poplar extrafloral nectaries: two types, two strategies of indirect defenses against herbivores. *Plant Physiol.* **159**, 1176–1191 (2012).
- Heil, M., Rattke, J. & Boland, W. Postsecretory hydrolysis of nectar sucrose and specialization in ant/plant mutualism. *Science* **308**, 560–563 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to D. Ehrhardt, H. Cartwright, J. Lindeboom, K. Barton and T. Liu for help with confocal and scanning electron microscopy. We thank D. Ehrhardt for providing specific endomembrane markers and M. Jia for conducting nectar sugar assays. We thank J. Danielson for help with phylogenetic analyses. This work was made possible by support from the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences at the US Department of Energy under grant number DE-FG02-04ER15542 to W.B.F. I.W.L. was supported by the fellowship of Department of Biology, Stanford University and Carnegie. X.-Q.Q. was supported by the Carnegie Institution and Scholarship Program of the Chinese Scholarship Council (2009635108). C.J.C.'s work was supported by a grant from the US National Science Foundation (#0820730). I.T.B. was supported by European Research Council advanced grant ClockworkGreen (293926) and the Max Planck Gesellschaft, and thanks C. Diezel for technical assistance.

**Author Contributions** I.W.L., L.-Q.C., C.J.C., I.T.B. and W.B.F. conceived and designed experiments. I.W.L., L.-Q.C., X.-Q.Q., D.S., B.-H.H., K.G., S.-G.K., D.K., P.M.K. and M.K.G. performed experiments. I.W.L., W.B.F., L.-Q.C., X.-Q.Q., D.S., B.-H.H., S.O., P.K., C.J.C. and A.R.F. analysed the data. I.W.L. and W.B.F. wrote the manuscript and I.T.B. and C.J.C. revised it.

**Author Information** The sequence for a full-length cDNA for BrSWEET9 is available in the NCBI Reference Sequence database under accession number KC790460. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.B.F. ([wfrommer@stanford.edu](mailto:wfrommer@stanford.edu)).

# Structure of a type IV secretion system

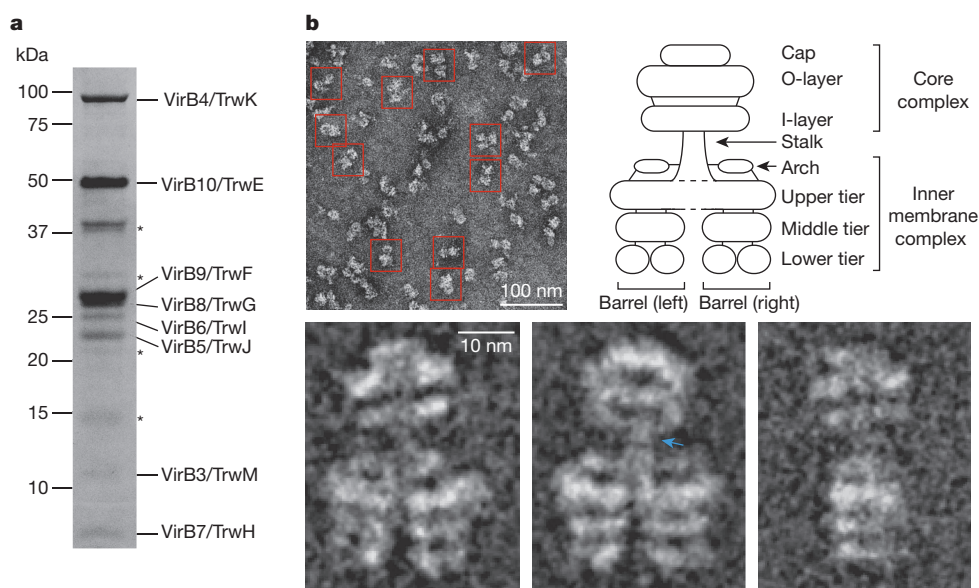
Harry H. Low<sup>1\*†</sup>, Francesca Gubellini<sup>2\*</sup>, Angel Rivera-Calzada<sup>1</sup>, Nathalie Braun<sup>1</sup>, Sarah Connery<sup>1</sup>, Annick Dujeancourt<sup>2</sup>, Fang Lu<sup>1</sup>, Adam Redzej<sup>1</sup>, Rémi Fronzes<sup>2</sup>, Elena V. Orlova<sup>1</sup> & Gabriel Waksman<sup>1</sup>

**Bacterial type IV secretion systems translocate virulence factors into eukaryotic cells<sup>1,2</sup>, distribute genetic material between bacteria and have shown potential as a tool for the genetic modification of human cells<sup>3</sup>. Given the complex choreography of the substrate through the secretion apparatus<sup>4</sup>, the molecular mechanism of the type IV secretion system has proved difficult to dissect in the absence of structural data for the entire machinery. Here we use electron microscopy to reconstruct the type IV secretion system encoded by the *Escherichia coli* R388 conjugative plasmid. We show that eight proteins assemble in an intricate stoichiometric relationship to form an approximately 3 megadalton nanomachine that spans the entire cell envelope. The structure comprises an outer membrane-associated core complex<sup>1</sup> connected by a central stalk to a substantial inner membrane complex that is dominated by a battery of 12 VirB4 ATPase subunits organized as side-by-side hexameric barrels. Our results show a secretion system with markedly different architecture, and consequently mechanism, to other known bacterial secretion systems<sup>1,4–6</sup>.**

The canonical type IV secretion (T4S) system comprises 12 proteins, VirB1–11 and VirD4, and forms a large macromolecular complex that spans the cell envelope of Gram-negative bacteria<sup>2</sup>. The hub protein VirB10 inserts into both the inner and outer membranes and spans the entire width of the periplasm. It is decorated by VirB7 and VirB9 in a 1:1:1 ratio to form a C<sub>14</sub> symmetrized outer membrane pore termed the core complex<sup>7</sup>. The architecture and relative topological positioning of the rest of the T4S system components, particularly at the inner membrane, are unknown. Three ATPases, VirB4, VirB11 and VirD4, energize the secretion apparatus<sup>1,2,8</sup>. Some T4S systems are known to extend a substantial tubular pilus, comprising VirB2 supplemented with a minor component, VirB5, beyond the cell boundary<sup>1,2</sup>.

To provide structural insights into an assembled T4S system, the *virB/trw* operon of the R388 plasmid from *virB1/trwN* to *virB10/trwE* (T4SS<sub>1–10</sub>) was cloned, overexpressed in *E. coli* and the solubilized membrane fraction subjected to affinity chromatography (Extended Data Fig. 1a). Analysis of eluted sample by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) (Fig. 1a), liquid chromatography mass spectrometry (LC–MS) and western blot (Extended Data Fig. 1b) reproducibly showed a purified complex consisting of VirB3/TrwM, VirB4/TrwK, VirB5/TrwJ, VirB6/TrwI, VirB7/TrwH, VirB8/TrwG, VirB9/TrwF and VirB10/TrwE (termed the T4SS<sub>3–10</sub> complex). Overexpression of *virB1/trwN* to *virB11/trwD* with *virB10/trwE* tagged as in the T4SS<sub>1–10</sub> clone yielded the same T4SS<sub>3–10</sub> complex.

To characterize the three-dimensional architecture of the T4SS<sub>3–10</sub> complex, a negative-stain electron microscopy data set was collected. Characteristic views of the complex (class averages) were obtained that typically showed double barrel-like densities connected by a thin central stalk to a structure clearly reminiscent of the pKM101 plasmid core complex<sup>7</sup> (Fig. 1b). Indeed, separate purification of R388 VirB7/TrwH, VirB9/TrwF and VirB10/TrwE, and negative-stain electron microscopy of the resultant complex, confirmed formation of a pKM101-like core complex complete with 14-fold symmetry (Extended Data Fig. 2). In the T4SS<sub>3–10</sub> assembly, those densities outside the obvious core complex are broadly termed the inner membrane complex (IMC) (Fig. 1b). Significant flexibility between IMC and the core complex was observed (Extended Data Fig. 3a–c), which prompted two discrete three-dimensional reconstructions to be generated for the two complexes (Extended Data Fig. 3d). Final resolutions of 18 Å and 23 Å were achieved, respectively (Extended Data Fig. 4a), with flexibility in the IMC accounting for the difference in resolutions. Ultimately, these reconstructions were merged

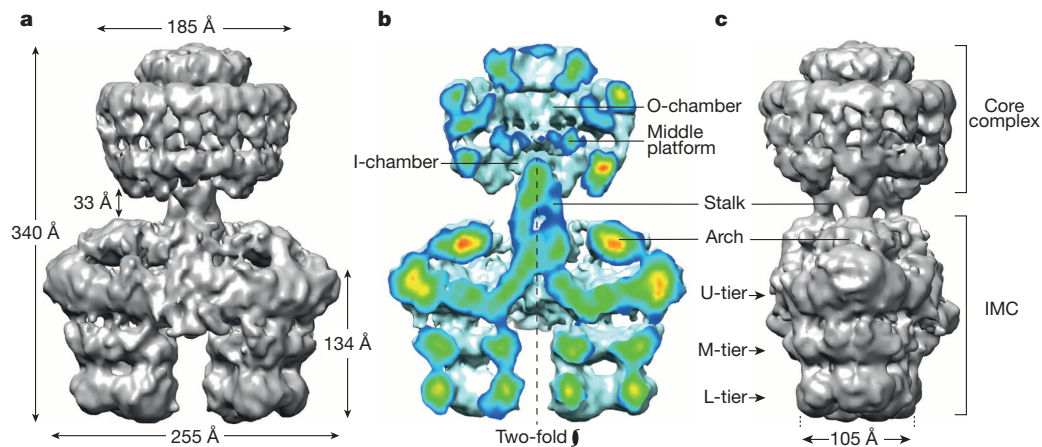


**Figure 1 | Purification of the R388 encoded T4SS<sub>3–10</sub> complex.** **a**, SDS–PAGE of the T4SS<sub>3–10</sub> complex. Asterisks indicate minor contaminants (from top to bottom: OmpF/OmpA, dihydrolipoyl dehydrogenase, single-stranded DNA-binding protein and lysozyme). **b**, Overview negative-stain electron microscope image and representative characteristic views (class averages) of the T4SS<sub>3–10</sub> complex with a schematic describing the nomenclature of observed structure components. Blue arrow indicates region of high flexibility.

<sup>1</sup>Institute of Structural and Molecular Biology, University College London and Birkbeck College, Malet Street, London WC1E 7HX, UK. <sup>2</sup>Institut Pasteur, G5 Biologie structurale de la sécrétion bactérienne and UMR 3528-CNRS, 25 rue du Docteur Roux, 75015 Paris, France. <sup>†</sup>Present address: Faculty of Natural Sciences, Imperial College, London SW7 2AZ, UK.

\*These authors contributed equally to this work.





**Figure 2 | Asymmetric composite structure of the T4SS<sub>3-10</sub> complex.** **a**, Front view. The map is a composite generated by merging independently processed core complex and IMC reconstructions. **b**, Cut-away front view. Electron density is colour-coded, ranging from red to blue, indicating regions of strong to weak density, respectively. The IMC has pseudo-two-fold symmetry around the particle long axis. **c**, Side view. U-, M- and L-tier substitute for upper, middle and lower tier, respectively.

to generate a composite map of the entire T4SS<sub>3-10</sub> complex (Figs 2 and 3a and Extended Data Fig. 4b–e). Final maps were generated with no symmetry applied. However,  $C_{14}$  symmetry is clearly observed in the core complex and overall  $C_2$  symmetry in the IMC (Fig. 2).

The whole T4SS<sub>3-10</sub> complex is 340 Å in length and comprises a core complex connected by a central stalk to an IMC that is 255 Å at its widest (Fig. 2). The observed core complex is similar to that of the well-characterized pKM101 plasmid<sup>9</sup>, with a diameter of 185 Å and rotationally symmetric inner (I-) and outer (O-) layers integrated to form a chamber that is divided by a central middle platform (Fig. 3b). The core complex cap is in an overall closed conformation as in the cryo-electron microscopy reconstruction of the pKM101 core complex<sup>9</sup>, with any central constriction<sup>7</sup> probably occluded by the negative stain of bound detergent or lipid, or both.

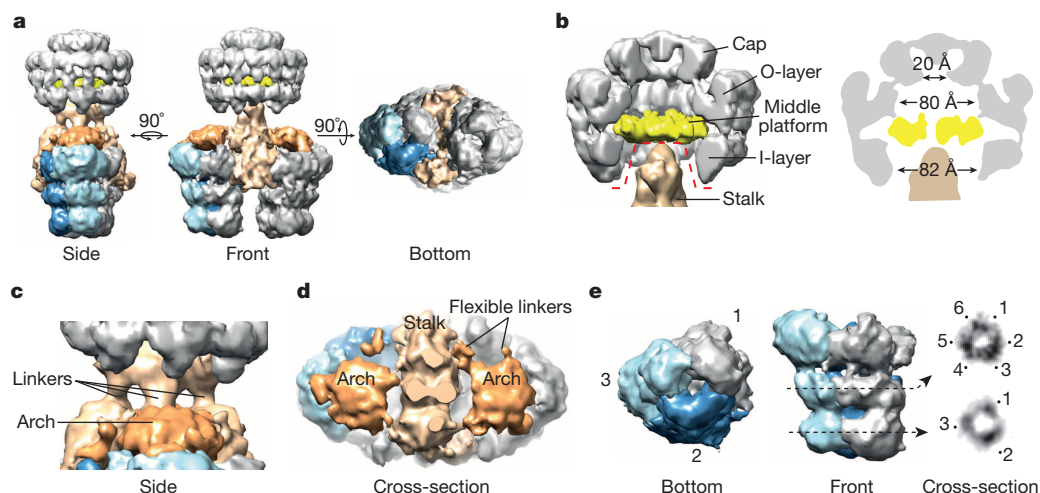
In contrast to previous T4S system core-complex reconstructions<sup>7,9,10</sup>, the inner (I-) chamber is partly occluded by the central stalk (Fig. 3b–d), which extends distally from the core complex to form a central binding hub within the IMC. The stalk does not show any clear symmetry, and in the composite T4SS<sub>3-10</sub> complex model it makes several limited contacts with the walls of the I-chamber. The top and middle parts of the stalk are linked by four connections that varied in intensity during refinement, with the central pair being most stable (Fig. 3c and Extended Data Fig. 5a). Other connections may exist that were not observed owing to marked flexibility in this region (Extended Data Fig. 3b, c).

Connected to the stalk distal end are two barrel-like densities related by  $C_2$  symmetry with length 134 Å and minimum diameter 105 Å (Fig. 2). Each barrel comprises three tiers, with the lower and middle tiers constituting rings with three-fold symmetry oriented around a central channel. In addition, a dimeric arrangement of densities around each three-fold axis is observed in the middle tier (Fig. 3e). The upper tier is more substantial and any central channel appears closed. Overall, each barrel comprises three basic elongated segments that are each made up of two subunits (Extended Data Fig. 5b).

Directly above the upper tier of each barrel lies a remarkable structure termed the arch, which interconnects between barrel subunits and the central stalk (Figs 2 and 3d). Each arch comprises a substantial central density that contacts the underlying barrel through up to six thin linkers (Extended Data Fig. 5a). Note that not all linkers were consistently resolved during refinement and often only a dominant triad of links was observed (positions 1, 3 and 5 in Extended Data Fig. 5a).

To ensure that dry preparation of samples did not affect the integrity of the particles, a tilt series experiment using cryo-negative-stain electron microscopy was performed. It confirmed the features and dimensions of the T4SS<sub>3-10</sub> complex reconstruction described above (Extended Data Fig. 6).

To begin dissecting the internal organization of the T4SS<sub>3-10</sub> complex, the stoichiometry of its constituents was determined using radioiodination<sup>11</sup>. A T4SS<sub>3-10</sub> complex clone variant was used in which a



**Figure 3 | Segmentation of the T4SS<sub>3-10</sub> complex reconstruction.** **a**, Side, front and bottom views. Of the two barrels, only the left one is segmented. The colour scheme used is upheld in all panels. **b**, Zoom cut-away view of the core complex and stalk. Dashed red line delineates the border at which the separate core complex and IMC reconstructions were merged (left). Central cross-section schematic of the core complex from this study (right). **c**, Zoom side view

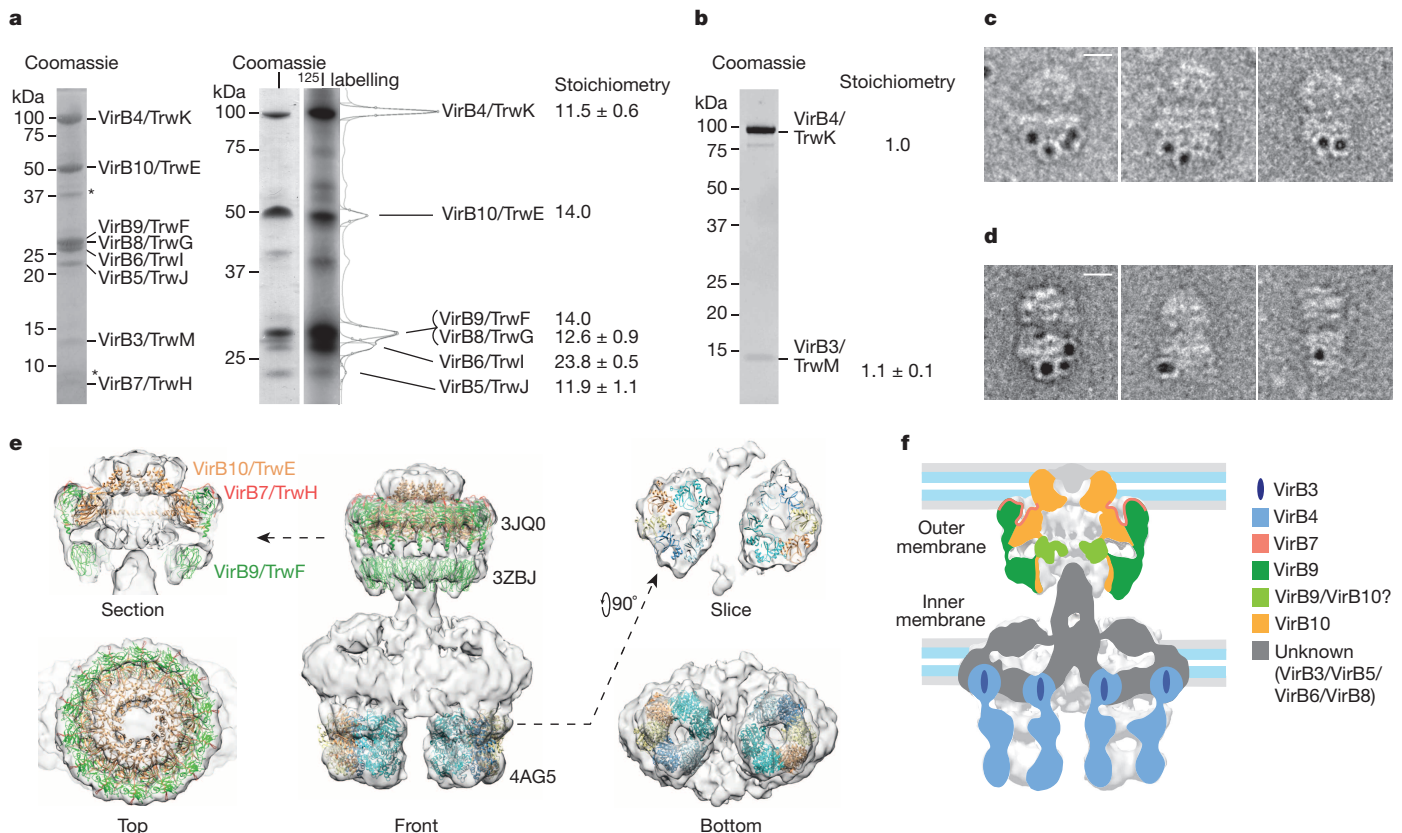
of the stalk. Some of the linkers between the core complex and IMC are flexible and were therefore poorly resolved. **d**, Cut-away top view of the stalk and arches. **e**, Each barrel-like density consists of three dimeric elongated segments. Cross-sections of the lower and middle tiers (right panel) show three-fold symmetry with a trimer of dimeric densities present in the middle tier.

His-tag was incorporated at the amino (N) terminus of VirB6/TrwI (T4SS<sub>3-10</sub>/His6-B6). The T4SS<sub>3-10</sub>/His6-B6 complex showed improved biochemical stability and was equivalent to the T4SS<sub>3-10</sub> complex when analysed by electron microscopy (Extended Data Fig. 7a). Using the known 14-fold copy number of VirB10/TrwE<sup>7,12</sup> for calibration, the relative stoichiometry of VirB4/TrwK, VirB6/TrwI and VirB5/TrwJ is 11.5:23.8:11.9 (Fig. 4a). Deconvolution of overlapping iodine-125 signals for VirB9/TrwF and VirB8/TrwG, based on a copy number of 14 for VirB9/TrwF<sup>7,12</sup>, suggests there are 12.6 copies of VirB8/TrwG (Fig. 4a). VirB7/TrwH has no tyrosine residue but its stoichiometry within the core complex is known and is the same as VirB10/TrwE. As radio-iodination was unsuitable for VirB3/TrwM owing to a low iodine-125 signal, VirB4/TrwK and VirB3/TrwM were co-expressed independently and purified as a distinct complex from the membrane fraction. Gel filtration and SDS-PAGE showed that these proteins co-elute (Fig. 4b and Extended Data Fig. 7b). Quantification of the respective bands showed a stoichiometry of 1:1 (Fig. 4b). Strikingly, it therefore seems that all constituents of the T4SS<sub>3-10</sub> complex, excluding those of the core complex, essentially exist in multiples of 12. The combined stoichiometry data suggest the T4SS<sub>3-10</sub> complex is approximately 3.4 MDa, which correlates well with the calculated mass estimation derived from the map volume (~3.0 MDa).

The stacked three-tier architecture of the IMC barrels is reminiscent of the hexameric electron microscopy reconstruction of VirB4/TrwK<sup>13</sup>. Indeed, gold labelling of the T4SS<sub>3-10</sub> complex with a His-tag incorporated

at the carboxy (C) terminus of VirB4/TrwK (T4SS<sub>3-10</sub>/B4-His6) shows localization of label to both IMC barrels (Fig. 4c and Extended Data Fig. 8a, c). Negative-stain electron microscopy of VirB4/TrwK purified from the membrane fraction or as an N-terminal maltose-binding protein (MBP) fusion shows an elongated and flexible monomer with approximate length 105 Å (Extended Data Fig. 7c–e), which is consistent with the internal length of an IMC barrel. Such data, combined with the VirB4/TrwK stoichiometry and the observed trimer of dimer symmetry within each IMC barrel, support a model in which each barrel contains six VirB4/TrwK monomers. Gold labelling of VirB6/TrwI (T4SS<sub>3-10</sub>/His6-B6 complex) shows a similar localization pattern to that of VirB4/TrwK, which suggests the VirB4/TrwK C terminus and VirB6/TrwI N terminus are in the same region within the complex (Fig. 4d and Extended Data Fig. 8b, c).

To deduce the position of the inner membrane, the relationship between VirB4/TrwK and the inner membrane was investigated. During T4SS<sub>3-10</sub>/B4-His6 complex purification, sequential washing of the membrane fraction with 6 M urea, high pH and high salt solutions failed to separate VirB4/TrwK (Extended Data Fig. 7f), which suggests robust insertion of VirB4/TrwK within the inner membrane. Incorporating data from previous studies<sup>13,14</sup>, it is proposed that the lower and middle tiers of each VirB4/TrwK barrel house the cytoplasmic VirB4/TrwK C terminus, whereas the upper tier is either partly or wholly inserted within the inner membrane and houses the VirB4/TrwK N terminus. Importantly, the soluble VirB4 C-terminal ATPase domain from



**Figure 4 | Stoichiometric analysis and localization of various macromolecular components within the T4SS<sub>3-10</sub> complex reconstruction.** **a**, Iodine-125 labelling of T4SS<sub>3-10</sub>/His6-B6 complex constituent proteins. Left, SDS-PAGE of the T4SS<sub>3-10</sub>/His6-B6 complex. \* indicates minor contaminants (OmpF/OmpA and Lpp). Right, SDS-PAGE of <sup>125</sup>I-labelled proteins in left lane (Coomassie) and corresponding radiograph in right lane (iodine-125 labelling). Relative stoichiometry was calculated by integration of band intensity and is shown at right. Reported means and corresponding standard deviations are from two separate labelling experiments on four independent purifications. **b**, SDS-PAGE of the VirB4/TrwK and VirB3/TrwM complex

stained with SYPRO Ruby. Relative stoichiometry was calculated by integration of band intensity and is shown at right. Reported means and corresponding standard deviations are from two independent purifications. **c**, Gold labelling (5 nm) of VirB4/TrwK clusters around the IMC barrels. Scale bar, 10 nm. **d**, Gold labelling (5 nm) of VirB6/TrwI shows a similar localization pattern to that of VirB4/TrwK (see c). Scale bar, 10 nm. **e**, Fit of the VirB4 ATPase domain from *T. pseudethanolicus* (4AG5), the pKM101 outer membrane complex (3JQ0) and the *in silico* model of VirB9/TraO (3ZBJ). **f**, Summary schematic showing the localization of known components and position of cell membranes. For clarity, only VirB nomenclature is used in the colour key.



*Thermoanaerobacter pseudethanolicus* fits optimally as a trimer of dimer arrangement within the lower and middle tiers (Fig. 4e)<sup>10</sup>.

The compelling fit of the pKM101 outer membrane complex structure<sup>12</sup> and *in silico* VirB9/TraO I-layer model<sup>9</sup> into the core complex of the T4SS<sub>3-10</sub> reconstruction (Fig. 4e) highlights the unexpected distance (~40 Å) between the I-layer base and the proposed position of the inner membrane (Fig. 4f). The VirB10/TrwE N termini, extending distally from the base of the I-layer, either as part of the stalk or as distinct strands, must span this space to insert into the inner membrane (Extended Data Fig. 9a). This architecture suggests that substrate could also access the core complex secretion chamber from within the periplasm. Such a model is compatible with some T4S systems thought to use a two-step process<sup>15,16</sup>. Substrate would first cross the inner membrane using either the VirD4/TrwB coupling protein, VirB4/TrwK, or the Sec machinery in some other T4S systems. Then, secondarily, substrate would be channelled into the core complex and extruded (Extended Data Fig. 9b). Evidence is also mounting for interaction between VirD4/TrwB, VirB4/TrwK and VirB11/TrwD<sup>17</sup>. VirB4/TrwK hexamers could recruit, or even be replaced by, VirB11/TrwD or VirD4/TrwB, which would allow different enzymatic and functional gearing of the T4S machinery during the secretion cycle.

When compared with the type III secretion (T3S) system<sup>18</sup> (the only known structure for an assembled secretion system; Extended Data Fig. 10), the T4S system shows a fundamentally different architecture. While T3S systems are organized as a series of integrated ring-like structures that form a continuous tubular conduit, T4S systems show a more modular design with a two-fold symmetric inner-membrane complex conjoined by a central stalk to a concentric outer membrane channel. Thus T4S systems represent a radically different evolutionary approach, design and mechanism for the translocation of substrate across the bacterial cell envelope.

## METHODS SUMMARY

The region from *virB11/trwN* to *virB10/trwE* (T4SS<sub>1-10</sub>) was cloned from the *E. coli* R388 plasmid into pASK3c and pBAD vectors with a StrepII tag encoding sequence at the 3' terminus of *virB10/trwE*. Constructs T4SS<sub>1-10/B4-His6</sub> and T4SS<sub>1-10/His6-B6</sub> derived from pBAD:T4SS<sub>1-10</sub> had His6-tags inserted at the C and N termini of VirB4/TrwK and VirB6/TrwL, respectively. The pASK3c:T4SS<sub>3-4</sub> construct was used to express the VirB3/TrwM–VirB4/TrwK complex. The T4SS<sub>3-10</sub> complex was extracted from bacterial membranes by mild solubilization using a detergent mixture of 0.5% w/v *n*-dodecyl-β-D-maltopyranoside (DDM), 0.75% w/v Decyl Maltose Neopentyl Glycol (DM-NPG) and 0.1% w/v digitonin, and purified by the StrepII tag. Gold labelling of the T4SS<sub>3-10/B4-His6</sub> and T4SS<sub>3-10/His6-B6</sub> complexes used 5 nm diameter Ni-NTA nanogold beads. Stoichiometry was determined by iodine-125 incorporation in tyrosine residues using purified T4SS<sub>3-10/His6-B6</sub>. Samples were stained with uranyl acetate for negative-stain electron microscopy, or ammonium molybdate for cryo-negative-stain electron microscopy, on carbon-coated copper grids. Grids were examined using mostly a Tecnai F20 FEG microscope operating at a voltage of 200 kV. Images were recorded on a Gatan UltraScan 4000 CCD camera. For the T4SS<sub>3-10</sub> complex, a total of 21,680 particles were selected manually using BOXER<sup>19</sup>. Contrast transfer function (CTF) was estimated and corrected using CTFFIND3 (ref. 20) and BSOFT<sup>21</sup>, respectively. Particle alignment steps, multivariate statistical analysis (MSA)<sup>22</sup>, angular reconstitution, three-dimensional reconstruction and structure refinement were performed using IMAGIC<sup>23</sup> and SPIDER<sup>24</sup>. The core complex/stalk and IMC regions were then separately processed to generate two discrete maps. The two map densities were adjusted to equivalent standard deviation and then overlaid using the core complex I-layer as reference for superimposition.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 September 2013; accepted 27 January 2014.**

**Published online 9 March 2014.**

1. Fronzes, R., Christie, P. J. & Waksman, G. The structural biology of type IV secretion systems. *Nature Rev. Microbiol.* **7**, 703–714 (2009).

2. Schröder, G. & Lanka, E. The mating pair formation system of conjugative plasmids—a versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54**, 1–25 (2005).
3. Schröder, G., Schuelein, R., Quebatte, M. & Dehio, C. Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen *Bartonella henselae*. *Proc. Natl Acad. Sci. USA* **108**, 14643–14648 (2011).
4. Cascales, E. & Christie, P. J. Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* **304**, 1170–1173 (2004).
5. Johnson, T. L., Abendroth, J., Hol, W. G. & Sandkvist, M. Type II secretion: from structure to function. *FEMS Microbiol. Lett.* **255**, 175–186 (2006).
6. Marlovits, T. C. *et al.* Structural insights into the assembly of the type III secretion needle complex. *Science* **306**, 1040–1042 (2004).
7. Fronzes, R. *et al.* Structure of a type IV secretion system core complex. *Science* **323**, 266–268 (2009).
8. Kerr, J. E. & Christie, P. J. Evidence for VirB4-mediated dislocation of membrane-integrated VirB2 pilin during biogenesis of the *Agrobacterium* VirB/VirD4 type IV secretion system. *J. Bacteriol.* **192**, 4923–4934 (2010).
9. Rivera-Calzada, A. *et al.* Structure of a bacterial type IV secretion core complex at subnanometre resolution. *EMBO J.* **32**, 1195–1204 (2013).
10. Wallden, K. *et al.* Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proc. Natl Acad. Sci. USA* **109**, 11348–11353 (2012).
11. Passmore, L. A. *et al.* Structural analysis of the anaphase-promoting complex reveals multiple active sites and insights into polyubiquitylation. *Mol. Cell* **20**, 855–866 (2005).
12. Chandran, V. *et al.* Structure of the outer membrane complex of a type IV secretion system. *Nature* **462**, 1011–1015 (2009).
13. Peña, A. *et al.* The hexameric structure of a conjugative VirB4 protein ATPase provides new insights for a functional and phylogenetic relationship with DNA translocases. *J. Biol. Chem.* **287**, 29925–29932 (2012).
14. Durand, E., Oomen, C. & Waksman, G. Biochemical dissection of the ATPase TraB, the VirB4 homologue of the *Escherichia coli* pKM101 conjugation machinery. *J. Bacteriol.* **192**, 2315–2323 (2010).
15. Locht, C., Coutte, L. & Mielcarek, N. The ins and outs of pertussis toxin. *FEBS J.* **278**, 4668–4682 (2011).
16. Pantoja, M., Chen, L. S., Chen, Y. C. & Nester, E. W. *Agrobacterium* type IV secretion is a two-step process in which export substrates associate with the virulence protein VirJ in the periplasm. *Mol. Microbiol.* **45**, 1325–1335 (2002).
17. Ripoll-Rozada, J., Zunzunegui, S., de la Cruz, F., Arechaga, I. & Cabezon, E. Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems. *J. Bacteriol.* **195**, 4195–4201 (2013).
18. Schraut, O. & Marlovits, T. C. Three-dimensional model of *Salmonella*'s needle complex at subnanometre resolution. *Science* **331**, 1192–1195 (2011).
19. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semi-automated software for high-resolution single-particle reconstructions. *J. Struct. Biol.* **128**, 82–97 (1999).
20. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
21. Heymann, J. B. Ssoft: Image and molecular processing in electron microscopy. *J. Struct. Biol.* **133**, 156–169 (2001).
22. van Heel, M. *et al.* Single-particle electron cryo-microscopy: towards atomic resolution. *Q. Rev. Biophys.* **33**, 307–369 (2000).
23. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
24. Frank, J. *et al.* SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190–199 (1996).

**Acknowledgements** This work was financed by grant 098302 from the Wellcome Trust to G.W. and by equipment grant 079605 from the Wellcome Trust. R.F. and A.D. were financed by Institut Pasteur and the CNRS. F.G. was the recipient of 'Bourse Roux' from Institut Pasteur. We thank G. Péhau-Arnaudet for support with the electron microscopes at Institut Pasteur.

**Author Contributions** H.H.L., F.G., R.F., E.V.O. and G.W. designed the experiments. H.H.L., F.G., S.C. and F.L. generated the clones. H.H.L. first purified the T4SS<sub>3-10</sub> complex, and collected and processed the negative-stain electron microscope data for that complex. F.G., assisted by A.D., purified the T4SS<sub>3-10</sub> complex and performed the stoichiometry, gold labelling and membrane wash experiments for that complex. A.R.-C. purified, collected electron microscope data and processed the TrwH/TrwF/TrwE complex. N.B., assisted by A.R., collected T4SS<sub>3-10</sub> complex cryo-negative-stain data. H.H.L. and A.R.-C. purified, collected electron microscope data and processed TrwK and MBP–TrwK. A.R.-C. and F.L. purified the TrwM/TrwK complex and determined its stoichiometry. H.H.L. drafted the paper, and H.H.L., F.G., R.F., E.V.O. and G.W. finalized it.

**Author Information** The electron microscope composite map of the T4SS<sub>3-10</sub> complex has been deposited in the Electron Microscopy Data Bank under accession number EMD-2567. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.W. ([g.waksman@mail.cryst.bbk.ac.uk](mailto:g.waksman@mail.cryst.bbk.ac.uk)), E.V.O. ([e.orlova@mail.cryst.bbk.ac.uk](mailto:e.orlova@mail.cryst.bbk.ac.uk)) or R.F. ([remi.fronzes@pasteur.fr](mailto:remi.fronzes@pasteur.fr)).



## CORRIGENDUM

doi:10.1038/nature13280

### **Corrigendum: Rapid remobilization of magmatic crystals kept in cold storage**

Kari M. Cooper & Adam J. R. Kent

*Nature* **506**, 480–483 (2014); doi:10.1038/nature12991

In Fig. 1 of this Letter, a drafting error meant that the labels for the vertical axis were offset by one order of magnitude (for example, the tick mark labelled '10<sup>3</sup>' should have been labelled '10<sup>4</sup>'). Figure 1 has been corrected online.

## CORRIGENDUM

doi:10.1038/nature13299

### **Corrigendum: C11orf95–RELA fusions drive oncogenic NF–κB signalling in ependymoma**

Matthew Parker, Kumarasamypet M. Mohankumar, Chandanamali Punchihewa, Ricardo Weinlich, James D. Dalton, Yongjin Li, Ryan Lee, Ruth G. Tatevossian, Timothy N. Phoenix, Radhika Thiruvengadam, Elsie White, Bo Tang, Wilda Orisme, Kirti Gupta, Michael Rusch, Xiang Chen, Yuxin Li, Panduka Nagahawhatte, Erin Hedlund, David Finkelstein, Gang Wu, Sheila Shurtleff, John Easton, Kristy Boggs, Donald Yergeau, Bhavin Vadodaria, Heather L. Mulder, Jared Becksfort, Pankaj Gupta, Robert Huether, Jing Ma, Guangchun Song, Amar Gajjar, Thomas Merchant, Frederick Boop, Amy A. Smith, Li Ding, Charles Lu, Kerri Ochoa, David Zhao, Robert S. Fulton, Lucinda L. Fulton, Elaine R. Mardis, Richard K. Wilson, James R. Downing, Douglas R. Green, Jinghui Zhang, David W. Ellison & Richard J. Gilbertson

*Nature* **506**, 451–455 (2014); doi:10.1038/nature13109

In this Article, author Jared Becksfort's surname was incorrectly spelled 'Becksford'. Also, his affiliation should be the Department of Computational Biology and Bioinformatics (not the Department of Pathology) at St Jude Children's Research Hospital. His name and affiliation have been corrected online.

# CAREERS

**TURNING POINT** Plasma physicist-turned-educator aims for public office **p.557**

**WORK-LIFE BALANCE** Study finds that burnout is more likely for women **p.557**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



## GRANTS

# Funder storm

*A confluence of budget disappointments has confounded US scientists and left many uncertain about their future.*

BY HELEN SHEN

Last May, Heidi Fisher feared that her career as an evolutionary geneticist might be over. Although she was a postdoc in a large, successful lab at Harvard University in Cambridge, Massachusetts, and had numerous publications under her belt, she was struggling to secure a faculty position at a university where

she could continue her studies on adaptive mating strategies in deer mice.

Her 2012 application for a K99 grant from the US National Institutes of Health (NIH) in Bethesda, Maryland — funding that would help her to establish an independent lab — scored well by a typical year's standards. But looming spending cuts in 2013 and a highly uncertain financial outlook pushed back the

NIH's payline (the score separating grants that will be funded from ones that fall short), bumping Fisher's application into a ballooning 'maybe' category, to be decided on once the agency could fully assess its budget prospects.

After a nail-biting, half-year delay, Fisher heard that she would not receive the grant. Crushed, she considered seeking a job in biotechnology, education, legal services or even art. "I thought, 'If I can't get a grant now, am I going to be struggling like this my whole career?'" she says.

Then, in an unexpected turn of events a month later, she learned that the NIH had found a small amount of extra funding to support her award. She has now been offered a faculty position at a leading US research university.

Such twists and turns have become more common across the United States as a result of drawn-out budget disruptions. First came across-the-board government spending cuts known as sequestration — the result of Congress's failure to agree on a deficit-reduction strategy — which lopped 5% from science agencies' budgets on 1 March 2013. Then, in October, partisan budget battles forced a 16-day government shutdown (see *Nature* **502**, 13–14; 2013) that suspended nearly all federally funded science programmes, with dire consequences for some time-sensitive research.

The outlook has brightened somewhat, however: in January, Congress released a 2014 spending deal that suspended sequestration for two years and slightly boosted the budgets for the US National Science Foundation (NSF) in Arlington, Virginia, and the US Department of Energy (DOE) Office of Science over 2012 levels. The NIH, however, took a US\$800-million cut, continuing a ten-year downward trend (see *Nature* **505**, 461–462; 2014).

But some observers worry that last year's trauma will continue to influence the attitudes and decisions of scientists in training for some time. "A graduate student who looks at their principal investigator and sees how much trouble they're having is discouraged from going out into academia and setting up their own lab," says Laurie Glimcher, dean of Weill Cornell Medical College in New York City. "I'm sure we're going to lose some of the next generation of scientists."

## GRIM STATISTICS

Sequestration slashed funding for the NIH by \$1.55 billion, and grant-application success rates fell to a historic low of 16.8%, according to Sally Rockey, NIH deputy director ►

AKINOVETTA/GETTY



► for extramural research. In an analysis last month of R-series grants — the most common type of NIH research grant — the American Society for Biochemistry and Molecular Biology in Rockville, Maryland, reported that the agency funded 1,000 fewer people last year, a 4% drop from the previous year: the number dropped by only 150, or 0.5%, between 2011 and 2012. “Not getting funded for one year is not necessarily the end of a career,” says the society’s president, Jeremy Berg, “but some people are at the end of their rope.”

Yuntao Wu, an HIV researcher at George Mason University in Manassas, Virginia, ran out of NIH funding in April 2013 after two failed attempts to renew his R01 grant. He laid off his only technician, and for the past year has been running his lab on about \$3,000 per month, scraped together from university funds, industry contributions and a private fundraising campaign. The budget barely covers the cost of keeping his cell cultures alive and conducting a few small experiments, says Wu. He submitted around 14 applications to the NIH last year, and is optimistic that at least one will be successful. But he worries about the lasting impressions of the past year on his graduate students, whose research has slowed to a crawl. “They saw me struggle. They saw the lab really struggle,” he says.

### BUMPY RIDE

NIH-funded researchers were not alone in suffering the sting of sequestration. At the NSF, 690 fewer grants were awarded in 2013 — down 6% from 2012 — according to a report released by the US Government Accountability Office last month.

The physical sciences have seen some recovery after the initial shock in March 2013. Sequestration had threatened to close or delay construction on several large-scale physics initiatives, including the Alcator C-Mod magnetic fusion experiment at the Massachusetts Institute of Technology in Cambridge (see *Nature* **487**, 420; 2012). But in June 2013, responding to Congress, the DOE shifted some money to sustain the C-Mod and cancel scheduled lay-offs (see *Nature* **498**, 527–528; 2013). The project, together with other physics initiatives, also received modest, life-saving funding boosts as part of the \$1.1-trillion 2014 federal spending deal approved by Congress in January (see *Nature* <http://doi.org/r73>; 2014).

Other big projects, such as the Alaska Volcano Observatory (AVO), have not fared so well (see *Nature* <http://doi.org/r74>; 2013). Funding for the University of Alaska Fairbanks’ part in the volcano-monitoring programme — a collaboration with the US Geological Survey in Reston, Virginia, and the Alaska Division of Geological and Geophysical Surveys in Fairbanks — shrank from \$1 million to \$513,000, according to Jeff Freymueller, coordinating scientist for the project. And that topped off years of dwindling research budgets, he says. “We’re

on the verge of losing monitoring capabilities on a couple of volcanoes.”

Losing real-time eruption data from the monitoring network has dealt a blow to the next generation of volcanologists, adds Freymueller, as well as threatening aviation safety.

Since the sequester cuts, the university has not been able to replace the outgoing AVO graduate students who had been mining data from the roughly 200 seismic stations for insights into how volcanoes work.

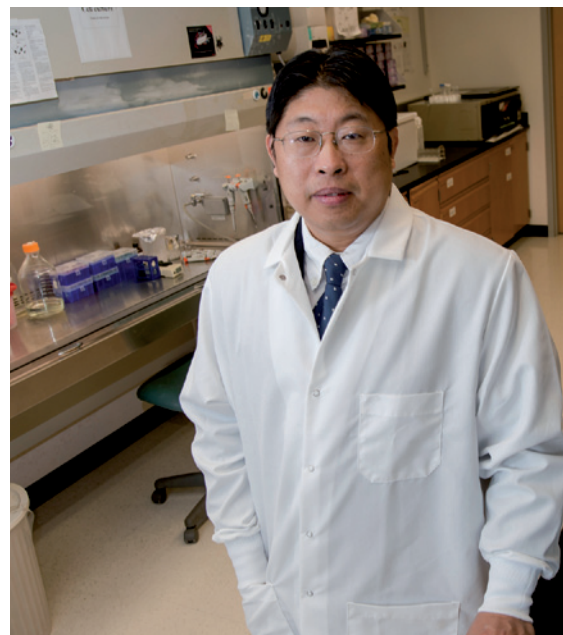
### TRAINING TEST

The full effects of sequestration on graduate students and postdocs around the country have been difficult to quantify. There are some regional indications, however: last October, a survey of 230 trainees at the University of California, San Francisco, showed that 71% of graduate students and 78% of postdocs thought that the current budget climate had decreased their desire to pursue research careers in academia. The university’s Science Policy Group, the campus organization that designed the survey, hopes to repeat the survey in other universities.

Career-development professionals at several institutions have noticed a few common patterns emerging among young scientists. For example, at the Jefferson Graduate School of Biomedical Sciences in Philadelphia, Pennsylvania, Lisa Kozlowski, associate dean for postdoctoral affairs and recruitment, is counselling more and more postdocs who have been given just a few months to get their affairs in order and leave their labs when their funding runs out.

The turmoil of the past year should serve as a wake-up call for all trainees, says Kozlowski, even for those who escaped the worst effects of the sequestration. She says that too many graduate students and postdocs procrastinate on making long-term career plans, or simply overlook their importance. Advance planning, including professional networking and cultivating leadership skills, are more important than ever in an uncertain funding climate. “Start early, even when you think you’re set in the lab and well-funded,” she advises.

Many career-development professionals say that bleak budgets are pushing trainees to look outside of academia. Sibby Anderson-Thompkins, director of postdoctoral affairs at the University of North Carolina at Chapel Hill, saw a particular shift last year in interest



Yuntao Wu was unable to renew his NIH grant.

towards careers in science communication and science policy. “I think they realize with sequestration that decision-makers often don’t understand research and the impact of research,” she says.

One of those mulling such a move is postdoc Melissa O’Connor of the Medical University of South Carolina in Charleston. Facing tight budgets, her supervisor elected to spend the rest of his R01 grant on collecting data to bolster a future grant application rather than applying to renew his funding. That means that support for O’Connor’s work on immune-cell signalling will run out in May.

O’Connor has dabbled in science advocacy, and in October 2012, she was selected for a one-year science-policy fellowship by the American Society for Pharmacology and Experimental Therapeutics in Bethesda. “Whether or not I get a career in that direction,” she says of her new-found interest, “it’s always going to be something I’m actively involved in.”

Soon enough, O’Connor and other young scientist-advocates may find themselves working to deal with a fresh set of budget challenges. As Congress negotiates a federal budget for the 2015 fiscal year, the approaching elections will probably polarize discussions along the same party divides that touched off last year’s chaos, notes Jennifer Zeitzer, director of legislative relations at the Federation of American Societies for Experimental Biology in Bethesda. And unless Congress can agree on an alternative plan, sequestration is set to restart in 2016. “It’s just a little bit of breathing room before the next round of budget problems,” says Zeitzer. ■

**Helen Shen** is a freelance writer based in Mountain View, California.

EVAN CANTWELL/GEORGE MASON UNIV.



**“If I can’t get a grant now, am I going to be struggling like this my whole career?”**

Heidi Fisher

# TURNING POINT

## Andrew Zwicker

*Andrew Zwicker, head of science education at the Princeton Plasma Physics Laboratory in New Jersey, announced his candidacy on 9 March as a Democratic congressman for the state. If elected, he'll be the fourth physicist ever to hold a federal legislative office.*

### How did you end up making science education such a big part of your career?

Two career-changing events happened when I was a postdoc. First, I met Rush Holt before he became US Representative Holt (Democrat, New Jersey). At the time, he was creating the Plasma Physics Laboratory's science-education programme, and recruited me as a mentor. Second, I mentored a student from a public high school in Trenton, New Jersey, who had amazing maturity and drive but was behind in her schoolwork. I spent hours helping her to write a computer program to analyse scientific data. She got a full scholarship to university, and I learned how rewarding it is to mentor young scientists. Soon after, I was named as head of the science-education programme. I went from fusion researcher to science educator. Now I'm researching the use of electric glow discharges, such as those from fluorescent lamps, as teaching aids.

### How have your interests evolved?

While leading the science-education programme, I was interested in ways to get people engaged in science. In 2006, I won first prize in an art-science competition run by the university for an image I took of a floating dust cloud of silica microspheres suspended in a plasma and illuminated by laser light (see [go.nature.com/ckpmnl](http://go.nature.com/ckpmnl)). But I am perhaps proudest of receiving an honourable mention for a video explaining what a flame is to an 11-year-old — an entry to a competition run by the Alan Alda Center for Communicating Science at Stony Brook University in New York. Scientists are stereotyped as being bad communicators, but I feel that we have a responsibility to communicate what we are doing with taxpayers' money. If the public isn't behind us, then appropriators in Congress will not see the value of our work.

### Have you always had political aspirations?

No. I can't say that running for Congress has been a lifelong ambition. When Rush announced that he wasn't going to run again, lots of people were caught off guard. Rush has a lot of respect in New Jersey. People started asking me if I would consider running because my narrative is so similar to his — we are both plasma physicists involved with



Princeton's science-education programme. When a Republican approached me, saying that the country needed people who use facts and evidence to make decisions rather than being motivated by other factors and ideologies, I started to seriously consider running.

### What advice has Holt offered you?

We have discussed the current state of affairs in Washington DC and the difficulties associated with gridlock. He told me that both Democrats and Republicans would turn to him or his staff on science issues. So being a scientist could help me to reach both parties.

### What persuaded you to run?

The undergraduates on campus were so enthusiastic, and asked hard questions. They desperately want the government to address big issues such as energy and climate and to help to make Earth a better place. I know I was talking to youthful, idealistic people, but it was overwhelming. I realized I had to do it.

### Does being a physicist help or hinder your chances?

There are four candidates. I would not be a viable candidate in this district if I were not a scientist — three of my opponents are career politicians, have been in some level of politics for 20 years and have name recognition. One of my goals is to get younger people, who do not usually vote in mid-term elections, to the polls.

### What's your campaign slogan?

Rush is well known for a bumper sticker that reads: "My Congressman is a rocket scientist." If people vote for me, they can keep their bumper sticker and just add the word "still". ■

INTERVIEW BY VIRGINIA GEWIN

## WORK-LIFE BALANCE

### Burnout predictors

Women are more likely than men to find that work interferes with their home life, and consequently are more likely to burn out, finds a study (V. Blom *et al. J. Occup. Environ. Med.* **56**, 361–366; 2014). The authors surveyed 4,446 twins in Sweden to investigate the determinants of work-home and home-work interference. Although genetic effects were minimal, they found differences in stress levels between the sexes. Lead author Victoria Blum, a psychologist at the Karolinska Institute in Stockholm, says that women typically have more home responsibilities and thus a greater total workload than men. To avoid burning out, she says, researchers need to find ways to work from home when possible and to share home-based responsibilities more evenly.

## IMMIGRATION

### UK visa fast-track

Research Councils UK (RCUK) is piloting a scheme to streamline the entry of researchers into Britain under the 'Tier 1 (Exceptional Talent)' visa. It has teamed up with the Royal Society, the British Academy and the Royal Academy of Engineering to guarantee endorsement for researchers who are awarded specific RCUK fellowships and grants, a process that usually takes nine weeks. Rosie Beales, senior policy manager at the RCUK, says that the agency will assess the pilot scheme's success in the next six months. Representatives of UK universities and scientific organizations have warned that complicated visa procedures have made Britain an unattractive destination for scholars (see *Nature* **506**, 14–15; 2014).

## SOCIAL MEDIA

### Science 2.0

Using social media more often would help scientists to disseminate their results, debate findings and engage a wider audience, says a study (C. Greenhow and B. Gleason *Br. J. Educ. Technol.* **45**, 392–402; 2014). Co-author Christine Greenhow at Michigan State University in East Lansing, says that researchers must learn to create a robust online presence. As part of an upcoming study, she polled 1,600 US and Canadian faculty members and found that 15% use Twitter, 28% use YouTube and 39% use Facebook for scholarly activity. Greenhow is calling for social-media metrics to be added to the tenure process.